Birth Spacing and Child Survival: Comparative Evidence from India and Pakistan*

Pushkar Maitra♣
Monash University

and

Sarmistha Pal♠
Brunel University

June 2005

Abstract

In view of higher fertility and mortality rates in Pakistan compared to India, this paper examines the two-way relationship between birth interval and child mortality and compares the behaviour of households in the Indian and Pakistani provinces of Punjab. Birth interval and child survival are modelled here as correlated hazard processes to address the bias generated by the simultaneity between spacing and survival. We find evidence of significant mutual dependence between birth interval and child survival in both samples. We also identify a close correspondence between birth interval and duration of breastfeeding and argue that the duration of breastfeeding is a good instrument of birth spacing in our samples. There are also interesting differences between Indian and Pakistani households with respect to effects of son preference and female literacy. We argue that part of these differences could be explained by differences in religion and state policies in these two neighbouring states.

Key words: Birth spacing, Child survival, Sibling competition and child replacement effects, Religion and state policy, Correlated hazards models, Simultaneity bias.

JEL classification: J13, O10, C41, C24

---

* Pushkar Maitra wishes to thank the Faculty Research Grant Scheme, Faculty of Business and Economics, Monash University for research support. Much of this research was done when Sarmistha Pal was based in Cardiff and she wishes to thank Cardiff University and Cardiff Business School for funding this research. We are particularly grateful to the Managing Editor and also two referees for very constructive comments. We would also like to thank John Cleland, Yvonne Dunlop, John Hunter, Brett Inder, Gerry Makepeace, participants at the Australasian Labour Econometrics Workshop held at the University of New South Wales and the Econometric Society North American Summer Meetings at Evanston for their comments and suggestions. The usual caveat applies.

♣ Pushkar Maitra, Department of Economics, Monash University, Clayton Campus, VIC 3800, Australia. Email: Pushkar.Maitra@BusEco.monash.edu.au.

♠ Corresponding author: Sarmistha Pal, Department of Economics and Finance, Brunel University, Uxbridge UB8 3PH, UK. Email: sarmistha.pal@brunel.ac.uk .

Birth Spacing and Child Survival: Comparative Evidence from India and Pakistan

1. Introduction

This paper examines the inter-relationship between fertility and child mortality using data from two adjoining provinces in India and Pakistan. Given the biological constraints on child bearing, increased duration between children implies lower lifetime fertility. Child mortality on the other hand is an important measure of "child quality" and increased child quality is an essential component of the overall process of development of a country – after all these children are the workforce of the future and anything that has the potential to reduce the quality of children should be a matter of grave concern to economists, demographers and policy makers.

Quantity-quality trade-off is central to an understanding of household consumption and fertility decisions (Becker and Lewis, 1973; Becker, 1981). On the one hand, resource constrained households care about current income and hence might choose to have more children. The decision to have more children is typically reflected in shorter duration between children. On the other hand, to the extent children continue to live with their parents as adults, children of higher quality are likely to contribute more to household resources in the future. Therefore if households maximise the net present value of lifetime earnings, they will prefer to have children of higher quality. An increase in the number of children and/or shorter spacing will reduce the health of the children (via reduced allocation of resources per child and also parental efforts to distribute resources equitably among living children) and their future earning capacities. This trade-off justifies our interest in examining whether there is an empirically significant relationship between birth spacing and child survival in our samples.

Accordingly we hypothesize a two-way relationship between fertility and child mortality. Parental investment in children crucially depends on the duration between successive births, especially if parents are resource constrained. In particular, the closer apart the children are (i.e., the shorter the age difference between successive children), the greater is the competition among siblings for limited parental care and resources and the greater is the potential of the child not surviving. This is known as the sibling competition effect. Shorter birth interval also means more maternal depletion and therefore lesser ability of mothers to take care of young children. Early child death on the other hand might also result in a reduction in the duration between successive children because parents want to replace children that have died. This is known as the child replacement effect.

Much of the existing empirical evidence on the relationship between fertility and child mortality is derived from the estimation of single child health function (for example measures of child survival, child mortality, anthropometrical indicators, like weight-for-age, height-for-weight) only. While this literature tends to ignore the possible simultaneity bias arising from the inclusion of various family composition variables in the estimation of indicators of child health, it generally highlights the role of income and poverty (Behrman and Knowles, 1999), parental, especially mother's, education (Behrman and Wolfe, 1984), as well as that of birth interval, birth order and sibling characteristics like number of brothers, sisters, number of older brothers, sisters (Dasgupta, 1997; Garg and Morduch, 1998; Pal, 1999) in developing countries. We however argue that not only is child health closely related to parental decision on spacing of consecutive births, parental decision on birth spacing might also be closely related to the health of existing children. Consequently, one cannot treat birth spacing to be exogenous while determining child mortality and vice versa.

In this paper we adopt a unique technique to address this bias generated by simultaneity between spacing and survival. In particular, we treat birth spacing and child survival as correlated hazard processes where the hazard of child mortality depends on the duration to the next birth and the hazard of subsequent birth depends on child survival controlling for other individual (child specific), parental/household and community characteristics that can potentially affect the hazard of having a subsequent birth and the hazard of child mortality. The novelty of our approach is that we allow for mother-level unobserved heterogeneity in both spacing and survival hazard equations and assume that these two heterogeneity terms are correlated since the same woman/couple makes these spacing and allocation (which in turn affects child mortality) decisions. In other words, this approach enables us to model the mother-specific unobserved heterogeneity as common mother-level fixed effect, thus removing the bias resulting from the correlation, in an attempt to obtain corrected estimates of spacing and mortality hazards. Consequently, we are able to account for the indirect (and bias-corrected) effect of fertility/spacing on child mortality, usually overlooked in the literature.

The analysis is based on the National Family Health Survey (NFHS) 1992 – 93 data from the Indian province of Punjab and the Demographic and Health Survey (DHS) 1991–92 data from the Punjab province in Pakistan. The comparison between India and Pakistan generates obvious interest: while households in these provinces on either side of the border share a common history, the institutional environments (primarily pertaining to religious and political institutions) they live in have evolved very differently in the two countries since 1947 (Indian independence and the birth of the Pakistan nation). While India remained a secular state since 1950, Pakistan became an Islamic state after 1977. There has also been a clear distinction in the attitude of the state towards population planning: while India was one of the first British colonies to launch its population policy as early as 1951, Pakistani state remained rather passive towards any official population programme until early 1990s. Given the common history of the two regions, choice of our samples could, in some way, allow us to identify and hence evaluate the effects of religious and political institutions on differential demographic trends in these two provinces.

The paper contributes to the literature in several ways. First, although in recent years there has been a renewed interest on the effects of high fertility (measured by shorter duration between births and also concentration of births) on levels of infant and child mortality (especially in Pakistan, see for example Cleland and Sathar, 1984), most existing models of mortality treat fertility to be purely exogenous. We are not aware of any existing study that jointly estimates birth spacing and child survival as correlated hazards in an attempt to redress the endogeneity bias of single equation child health estimates. Second, we identify breastfeeding as the behavioural/biological mechanism that affects the relationship between fertility and child mortality. Since information on breastfeeding is not available for all children in the full samples, analysis of breastfeeding is based on the children born in the last 3-5 years of the survey in the two samples. On the basis of non-parametric and parametric analyses, it appears that breastfeeding is a good instrument of birth spacing in both countries, which in turn strengthens our central result of mutual causation between birth spacing and child mortality obtained from the full sample. Finally, the comparison between Indian and

Pakistani Punjab provides a unique opportunity to examine the effects of religion and state policy on fertility and child mortality over the birth cohorts. This issue remains quite unexplored (and contentious) in the literature.[1] While these provinces are highly prosperous in their respective countries and share a common socio-cultural background, they are significantly different in terms of religious composition/institutions, state policies throughout the post-1947 period and especially since the Islamisation of Pakistan after 1977. Among other things, our results highlight the differential effects of female literacy and son preferences on spacing and mortality hazards in the two countries. We argue that the latter reiterates the differential role of religious and political institutions and the interaction between the two, if any, determining state policies in the two provinces over much of the post-independence period.

The rest of the paper is organised as follows. Section 2 rationalises the econometric methodology used to jointly estimate child survival and birth spacing. Section 3 discusses the data sets and selected descriptive statistics. Section 4 discusses the results. Section 5 examines the effect of breastfeeding on fertility and child mortality (and on the inter-relationship between the two). Finally Section 6 concludes.

2. Methodology

Much of the existing evidence of the relationship between fertility and mortality are based on individual (uncorrelated) estimates of the effect of child/infant mortality (assumed to be exogenous) on fertility and the effect of fertility (again assumed to be

[1]For example Caldwell (1986) has argued that Muslim societies are often predisposed to high fertility and unmet need for contraception (though the underlying rationale behind this observation has seldom been thoroughly investigated. One possible hypothesis is the lack of women's autonomy in the Islamic society though Jejeebhoy and Sathar (2001) reject this hypothesis for their comparative study on India and Pakistan.

exogenous) on child/infant mortality. In this paper we argue that it is important to treat fertility and mortality as jointly determined variables as it allows us to examine the nature of mutual causation between these two variables.

The two variables of interest in our analysis are the hazard of child mortality and the hazard of birth of the next child following the birth of a particular child. An individual (woman who has ever given birth) may be observed over the duration of one or more child births. From the time a child is born the individual is at risk of having another child and/or the child dying. Both these hazards are influenced by a number of time varying factors (calendar time, the age of the mother, the time following the birth of the child, death of a particular child), a set of exogenous and (potentially) endogenous co-variates. Of particular importance is the effect of birth spacing on the hazard of child mortality (the sibling competition/resource constraint effect) and the effect of child mortality on the hazard of having the next child (the child replacement effect).

To be more specific, the log hazard of duration following the birth of the i^{th} child $(i = 1, \ldots, k)$ born to the j^{th} woman, $(j = 1, \ldots, n)$ may be written as:

$$h_{ij}^{n}(t) = \beta_0 + \beta_1 T_1(t) + \beta_2 X_{1ij} + \lambda_j^n + \varepsilon_{ij}^n \qquad (1)$$

and the log hazard of survival equation for the i^{th} child born to the j^{th} woman may be written as:

$$h_{ij}^{s}(t) = \alpha_0 + \alpha_1 T_2(t) + \alpha_2 X_{2ij} + \lambda_j^s + \varepsilon_{ij}^s \qquad (2)$$

Here X_{1ij} and X_{2ij} denote the two sets of explanatory variables that affect the hazard of survival and the hazard of the next birth respectively. Included in X_{1ij} is the age at death of the child if (s)he is dead at the time of the survey and included in X_{2ij} is the duration to the next child if (s)he is not the last child.

The unexplained component of both the log hazard of survival and the log hazard of duration is broken up into a component that is purely random (ε_{ij}^{n} and ε_{ij}^{s} in the two equations) and a component that is common to all children born to the same mother (λ_{ij}^{n} and λ_{ij}^{s} in the two equations), capturing the mother level unobserved heterogeneity that accounts for the unobserved mother specific biological or health endowments (for example, health or genetic endowments of the mother) common to all children born to the same woman. The unobserved heterogeneity terms are assumed to be uncorrelated with other explanatory variables but the two unobserved heterogeneity components λ_{ij}^{n} and λ_{ij}^{s} can be correlated because the same woman/couple makes decisions regarding birth spacing and resource allocation. For example parents might have some private information regarding the health of the mother (unobserved to the researcher), which makes children born to her susceptible to some health condition that increases the chances of the child not surviving. But that might also make the mother choose a higher level of lifetime fertility. In particular we will assume joint normality of these residual terms in the two log hazard equations i.e.,

$$\begin{pmatrix} \lambda^{n} \\ \lambda^{s} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{n}^{2} & \sigma_{ns} \\ \sigma_{ns} & \sigma_{s}^{2} \end{pmatrix} \right) \tag{3}$$

Ignoring this mother level unobserved heterogeneity terms can result in biased estimates. All other residual variation is captured by ε_{ij}^{n} and ε_{ij}^{s}, with $\varepsilon_{ij}^{n} \sim IIDN(0,1)$ and $\varepsilon_{ij}^{s} \sim IIDN(0,1)$.

Finally $T_1(t)$ and $T_2(t)$ represent separate "clocks" of duration dependence of the hazards that determine the baseline hazard. They are essentially splines in time since the individual becomes at risk of the event – risk of dying or risk of having a younger sibling. Let us denote the time at which an individual enters the risk of an event by t_0 and we subdivide the duration $t-t_0$ into N_i+1 discrete periods, which sum to the calendar time but which allow the slope coefficients to differ within ranges of time separated by the N_i nodes μ_k. The spline variable for the k^th^ interval between μ_{k-1} and μ_k is given by

$$T_k(t) = \text{Max}[0, \text{Min}(t-\mu_{k-1}, \mu_k - \mu_{k-1})]$$

Then the baseline log hazard function is defined as a spline or a piecewise linear function and the log hazard of the event will have different slopes over the duration. So the baseline hazard functions can be written as:

$$\begin{aligned} \beta_1 T_1(t) &= \sum_{k=1}^{N_1+1} \beta_{1k} T_{1k}(t) \\ \alpha_1 T_2(t) &= \sum_{k=1}^{N_2+1} \alpha_{1k} T_{2k}(t) \end{aligned} \tag{4}$$

In other words, the baseline log hazard is the sum of the effects of the various sources of time dependence within the period of risk for an individual and the resulting log hazard equation is piecewise linear in time since the episode began.

What we have therefore is a set of two correlated hazard equations. Notice that the log hazard of child mortality does not directly affect the log hazard of birth spacing equation (and vice-versa). Accordingly we estimate equations (1) and (2) jointly as a system of equations with the errors correlated across the two equations. We assume that the unobserved factors, which partly determine birth spacing and child mortality, are correlated because the unobserved mother-specific heterogeneity, which affects birth spacing, also affects child mortality. This is a novel approach to address the problem of endogeneity bias arising from the inclusion of spacing in the mortality equation or mortality into the spacing equation. By modelling this aspect of the data generating process as a common mother level effect, we are able to remove the implicit bias resulting from the correlation. See Lillard (1993), Brien and Lillard (1994), Brien et al. (1999) and Gangadharan and Maitra (2003) for more on this approach.

Denote $L^n\left(\lambda^n\right)$ and $L^s\left(\lambda^s\right)$ to be the conditional likelihood function of child mortality and time to next birth respectively, we can write the joint marginal likelihood as:

$$\int_{\lambda^n}\int_{\lambda^s}\prod L^n\left(\lambda^n\right)\prod L^s\left(\lambda^s\right)f\left(\lambda^n,\lambda^s\right)d\lambda^n d\lambda^s$$

Here $f\left(\lambda^n,\lambda^s\right)$ is the joint distribution of the unobserved heterogeneity components specified in equation (3). The full specification model is estimated using Full Information Maximum Likelihood (FIML) method.

Child mortality is defined as the number of months the child was alive (if he/she is dead at the time of the survey) or the age of the child at the time of the survey (if he/she is alive at the time of the survey), in which case the observation is censored. Our focus is on child deaths, so we restrict ourselves to child mortality in the age group 0 – 5 years and the children dead at the time of the survey but who died after the age of 5 are also regarded as being censored. Birth spacing (or birth interval) is defined as the interval, measured in months, between the reported dates of birth, rather than the inter-conception interval. In the case of the last child, the observed duration is the age of the child at the time of the survey and the observation is censored. The fact that we use the reported birth interval and not the inter-conception interval could mean that that the measured birth intervals might be shorter on account of pre-mature births (Gribble, 1993).[2] A related problem arises from the incorrect reporting on child death which in turn could affect birth interval and therefore its effect on child mortality. This is particularly a problem for older women who had given birth long time ago. Our estimates however seem pretty robust especially with respect to the relationship between birth interval and mortality as is confirmed by the estimates for the recent birth in the last 3-5 years of the survey. This robustness of our results can partly be attributed to the inclusion of woman/couple specific unobserved heterogeneity terms.

A further potential problem of using observed birth intervals is that the measured intervals might be longer on account of miscarriages and stillbirths. Unfortunately we do not have reliable information of the extent of miscarriage or still birth for each conception (remember that this can only observed if the woman *ever* had any miscarriage/stillbirth and reports it truthfully) in the sample and hence, we are thus unable to assess the impact of this problem. It is worth noting that ignoring mis-carriages and stillbirths might lead to an underestimation of the mortality effects of reduced birth intervals if it is the case that women who have this problem also produce weaker live births. In this case incorrectly measured long intervals might be associated with higher child mortality. This bias is however likely to be small in our models once we account for mother level unobserved heterogeneity and the correlation between this component of the error terms in the two equations.

There are of course alternatives to using the correlated hazard model used in this paper. An obvious alternative modelling technique might be to use a probit model to

[2] This however does not seem to be a very important issue in our sample. This is because when we re-estimate the equations by dropping all mothers with at least one birth interval less than 9 months from the sample, the resultant estimates remain virtually similar to the ones reported here.

estimate child mortality (along with a hazard equation for birth spacing). Maitra and Pal (2004) using data from Bangladesh do so: in that paper child mortality is measured using an indicator variable that takes the value of 1 if the child has died before the age of 10 (and is dead at the time of the survey) and 0 otherwise. However there are two potential problems with a probit mortality equation. First, it does not use all available information: in particular it does not use the information on the number of months a child is alive if he/she is dead at the time of the survey. Second, it is difficult to account for censoring in this case – remember that in the absence of longitudinal data, we do not know the final health outcome of the child. Accordingly, we argue that a correlated hazard model is a superior method to obtain simultaneity bias corrected estimate of child mortality.[3]

3. Data, Descriptive Statistics and Explanatory Variables

The empirical analysis is based on two data sets collected around the same time: the NFHS 1992 – 93 data[4] from India and the DHS 1990 – 91 data from Pakistan. We restrict ourselves to households residing in the Punjab provinces in the two countries. While the two countries differ in terms of their religious and political institutions, households in these two provinces on either side of the border share a common socio-economic and linguistic background because of their common origin. While GDP per capita is higher in Pakistan, India performs better in terms of the demographic measures of well-being: the infant mortality rate, the crude birth rate and the total fertility rate are

[3] While it could be argued that this is a strong assumption, it should be noted that our central result is robust to different specifications (1) – (4), see Tables 2 and 3.

[4] The second NFHS undertaken in 1998-99 was designed to strengthen the database further and facilitate implementation and monitoring of population and health programmes in the country. Though some additional information (e.g., height and weight of all eligible women, blood test for women and children) was collected, the information that we use remained very similar. Our preliminary analysis also yielded similar results as reported here.

all lower in India and adult literacy rates are lower in Pakistan.[5] These differences may in part be accounted for by differences in religious beliefs, which in turn have shaped the official policies and programs pertaining to population, education and employment in the two countries in the post-independence period.

Among all Indian states, as of 1991 – 92, Punjab had the highest per capita net state domestic product. It also had the lowest poverty rate (head count ratio) for both rural and urban regions of India. However in terms of demographic indicators the performance of Punjab is nothing to write home about. In this respect it is interesting to compare the performance of Punjab with that of Kerala, which has achieved demographic indicators, comparable to many developed countries. In 1991 – 92, net state output per capita in Kerala was half of that of Punjab. However the infant mortality rate is Punjab was more than three times that of Kerala (57 per thousand live births in Punjab, compared to 17 per thousand live births in Kerala) and the total fertility rate in Punjab is close to double that of Kerala (3.1 compared to 1.8). Finally male and female literacy rates were also significantly higher in Kerala.

Among the four Pakistani provinces (Punjab, Sindh, North-West Frontier Province and Balochistan), Punjab is the most prosperous and the most densely populated: more than 56% of all Pakistanis resided in Punjab in 1990. In terms of the different demographic and socio-economic indicators, Punjab has performed better compared to the rest of Pakistan. The average number of years of education for Punjabi women is 1.34 years compared to 0.91 for women residing in the rest of Pakistan. The average number of years of education for Punjabi men is 4.16 years, again significantly

[5] In 1992, the infant mortality rate was 79 in India compared to 95 in Pakistan; the crude birth rate was 29 per 1000 in India compared to 40 in 1000 in Pakistan; and total fertility was 3.7 in India compared to 5.6 in Pakistan. Adult female literacy rates were 39% in India and 22% in Pakistan; while adult male literacy rates were 64% in India and 49% in Pakistan.

higher compared to the rest of the country (3.33 years). Average household income in Punjab is also significantly higher compared to the rest of Pakistan.[6] Of the four Pakistani provinces, Punjab has the highest prevalence levels (though the NWFP experienced the most rapid rise in contraceptive use in the early 1990s).

While all households in the Pakistani sample are Muslims, most households in the Indian sample are either Sikhs (58%) or Hindus (39%) and only 1.5% of all households in the Indian sample are Muslims. One can identify certain behavioural differences between Muslim and non-Muslim households in the Indian Punjab. For example, compared to non-Muslim households, significantly lower proportion of Muslim parents were literate and were using some contraception (contraception use was even lower among households in Pakistani Punjab). It is now well documented that Hindus and Muslims also differ significantly in terms of their attitudes to son preference in different parts of South Asia. For example, Mutharayppa et al. (1997) find that compared to Hindus, son preference is generally lower among Muslims in India except Jammu and Rajasthan. Arnold et al. (1998) however argue that son preference has a negative effect on contraceptive use in Muslim dominated Bangladesh, regardless of socioeconomic and demographic characteristics. Hussain et al. (2000) find sex of surviving children is strongly associated with subsequent fertility and contraceptive behaviour. Thus son preference in fertility/spacing even among Muslims in many parts of South Asia can generate an indirect but significant ‘son preference’ effect in child mortality, as the probability of child survival is closely linked to fertility/spacing

[6] These figures were obtained from the Pakistan Integrated Household data set (PIHS) conducted in 1991. As with the DHS data, this was also a nationally representative unit record data set.

through resource competition effect. Perhaps these factors, at least in part[7], could explain why number of children ever born and mortality rates are both significantly higher among Muslim households in the Indian Punjab.

There are a total of 2995 women in the Indian sample, who have given birth to a total of 8798 children. However, as high as 40% Indian women were sterilised at the time of the survey and therefore we exclude the youngest child of these sterilised women. This reduces the number of sample children to 7896 of whom 51% were boys. About 34% of these children were first born (which also includes the only children). Of this total number 7896, 680 (about 8.6%) children died before reaching age 10 years (an overwhelming majority 71% of these children died before they were one year old). Average age at death was 11.52 months while mean duration of spacing was 30.26 months for the Indian sample. While the gender difference in spacing was not significant $(z = 0.562; p = 0.574)$, that in survival was $(z = 2.887; p = 0.004)$. The Pakistani sample consists of 8814 children born to 1955 women. In this case too we exclude the youngest children of the sterilised women. The sample consists of 4502 (51.08%) boys and 4312 girls. There were very few twins (7 to be exact) in the Pakistani sample and these women were deleted from the final sample though we retain the twins in the Indian sample. Of the 8814 children 1179 (13.38%) died before reaching the age 10 years and an overwhelming majority of the children who died before the age of 10 (72.43%) died before their first birthday. The mean age at death of the children that have died is around 14 months and the average duration between births is around 28 months. There is however no gender difference in child mortality

[7] One could possibly identify other related factors, namely, lower female literacy/autonomy, lower age at marriage, higher rate of high-risk pregnancy due to lower acceptance of modern family planning methods contributing to generally higher child mortality among Muslims in south Asia.

$(z = 0.363; p = 0.7169)$, while that in spacing is weakly statistically significant $(z = 1.833; p = 0.0668)$.

Table 1 presents the descriptive statistics for selected demographic variables in the two samples, which in turn reflect the differential demographic trend in the two provinces with different types of religious and political institutions. In particular, total number of children ever born is higher while duration between births is lower in Pakistan and the latter is accompanied by lower parental literacy and higher child mortality rates. The full set of descriptive statistics are available on request.

Explanatory Variables:

The explanatory variables X_{1ij} and X_{2ij} included in regressions (1) and (2) consists of a set of child specific (variables particular to each birth), parent/household specific variables (variables common to all children born to the same woman or born in the same household) and community level variables that affect the hazard of child mortality and the hazard of having the next child.

The log hazard of duration to next birth equation includes the set of explanatory variables SURV (the time the child was alive before dying) to capture the child replacement effect while the log hazard of duration to next birth equations includes the duration to next birth (NEXT) to capture the sibling competition or the resource constraint effect.

We include a dummy for gender of the child (BOY) in both regressions. Although biologically girls are less likely to die in the first year of their life (compared to boys), one could expect higher mortality rates among girls and a greater duration to the subsequent birth following the birth of a male child particularly if resource constrained parents have pro male bias, as is observed in many parts of South Asia,. Therefore any evidence of a statistically significant gender differential on child mortality rates, particularly in favour of boys could be symptomatic of severe discrimination, in terms of resource allocation, against girls.

Composition of siblings may affect child health outcomes in many low-income countries indirectly via their effects on spacing or may even have a direct effect. Even if we assume that parents cannot choose gender of a child (i.e., gender is exogenous), gender of the first child may influence parents to strategically determine subsequent birth spacing, by updating their fertility preferences. Thus given the gender of the child (known only after the child is born), parental decision to have an additional child will depend on the expected child earnings net of costs of bringing up a child[8] as well as the randomness associated with having another child of the desired gender. Thus parents characterised by son preference are more likely to increase the duration between successive births not only if the current child is a boy than if it is a girl, but also depending on how many of their existing children are girls. Gender composition of elder siblings is captured by an indicator variable ALLPREVFEM, which takes the value of one if all the previous children born to the woman are girls. A priori we expect that this should reduce the duration between successive births as the parents might keep trying to get a son.[9] It is however likely that the extent of son-preference may be

[8] This assumption accounts for the male-female differences observed in many south Asian societies including India (e.g., see Rosenzweig and Schultz, 1982). Very often female job opportunities are rather limited and more importantly the female child leaves parents' household after marriage while the male child when adult earns to look after the retired parents.

[9] In an earlier version of the paper we had instead used the proportion of elder siblings at birth that are females (PFEMOLD). However as an anonymous referee has pointed out that given the fact that parents in the Indian subcontinent often want at least one son their behaviour could be quite different depending on how many of the existing children are girls. As we discuss below, we find strong evidence that parents in Pakistan (not so in India) prefer to have at least one son; this effect cannot be adequately captured by

different between Muslim and non-Muslim households (see discussion earlier in this section). Sibling composition may also affect child mortality more directly. In general, if parents are 'resource constrained' then having several young children to care for will lead to a reduction in average resources per child, and this may offset the extra resources devoted to the youngest child even if that child is male. This scenario will, however, change if parents have pro-male bias; in the latter case, if parents have a high proportion of daughters among the older sibling group, they may devote *more* resources to the youngest child, especially if the youngest child is a boy.

To capture the effect of significant state dependence in child mortality we include an indicator variable, ANYPREVD, which takes the value of one if any of the previous children born to the woman have died. How does this effect work? There could be two opposite and confounding effects. On the one hand there is a learning aspect – one could argue that women who have had any of their children die have learnt from that unfortunate event and is less likely to repeat the same mistake for later children. Death of previous children therefore should therefore have a negative effect on child mortality. On the other hand there could be the so called maternal depression effect, and this could be particularly important if the last child has died. Death of a child, not surprisingly, could affect the mental state of the mother and this could have a flow-on effect on the actual health of the child and also on the parental decision to have an additional child. The final effect depends on which of the two effects is stronger. We did re-estimate the equations with LCHDEAD (last child dead) and FIRSTDEAD (first child dead) as alternative explanatory variables. These results are available on request. Finally in the case of India we included a dummy for twin births (TWIN).

---

simply including PFEMOLD as an explanatory variable. We would like to thank the anonymous referee for suggesting this.

One can possibly raise question as to why we have not included prior birth spacing (the duration between child i and child $i-1$: PRIOR). Prior birth spacing could have statistically significant effects on both birth spacing and on child health – essentially the relationship works through the maternal depletion effect. Low duration between child i and child $i-1$ could mean that the mother might not have had enough time to recover (in terms of health) after the birth of the previous child. That could have adverse effects on the health of child i. Additionally in this case parents might choose to delay the birth of the next child. The problem here is that when we included PRIOR as an explanatory variable in the hazard of duration to next birth regression leads to convergence problems for both samples. Further, including PRIOR as an explanatory variable in the hazard of child mortality regression leads to convergence problems for the Indian sample, but not for the Pakistan sample. We believe that this problem could partly be explained by the close correlation between PRIOR and NEXT (after all the two series include the same numbers arranged differently). Thus realistically one would expect that the effects of PRIOR and NEXT would be rather similar. In fact this is confirmed for the Pakistani case (see Appendix Table A2) for which we could obtain mortality estimates including both PRIOR and NEXT.

Let us now turn to the parental/household characteristics. Three indicator variables are included to capture the effect of the age of the mother at the time of child birth: AGEM2, AGEM3 and AGEM4 (age of the mother at the time of child birth 21 – 25, 26 – 30 and greater than 30). The reference category is that the age of the mother is 20 or less at the time of child birth. It has been argued that there is typically a u-shaped relationship between the age of the mother at the time of child birth and child mortality (or equivalently an inverted u-shaped relationship between the age of the mother at the time of child birth and child health/quality). Biologically speaking, early or late childbearing may be detrimental to the health of the foetus because of impaired functioning of a woman's reproductive system. Likewise the mother's age at the time of birth of child i could affect the duration between child i and child $i+1$ and we expect that the higher is the age of the mother at the time of the birth of the i^{th} child the lower is the duration between child i and child $i+1$, should the woman have another child. Additionally for the Pakistani sample we also include three dummies for the age of the father at the time of the birth of the child. Given the life-cycle effects of age on income, the age of the father could be viewed as a proxy for the permanent income of the household.[10]

Second, we include indicator variables for the educational attainment of the mother and the father (in both regressions) and an indicator variable for contraceptive use in the log hazard of duration to next birth equation. We include two dummies for the highest education attained by the mother and the father: EDUCM1 and EDUCM2 (the highest education attained by the mother is some primary school and the highest education attained by the mother is completed primary school or higher respectively) and two dummies EDUCF1 and EDUCF2 (the highest education attained by the father is some primary school and the highest education attained by the father is completed primary school or higher respectively) in the case of Pakistan and a dummy LITDAD (the father can read and write) in the case of India. One would expect mortality rates to be lower for children born to educated parents. Education lowers the cost of information and it is likely that more educated parents have a better knowledge and understanding of

[10] Note that in the Indian case however father's age variable had a number of missing observations (and the information does not seem to be very reliable as it was provided by the woman) and none of the father's age dummies were ever statistically significant even when we tried to include them.

the health conditions and health services and facilities available are better able to utilize the facilities and services available. Educated parents are also likely to be more aware of (potentially) adverse health effects associated with reduced birth spacing. In addition the father's age and educational attainment variables could be viewed as proxies for permanent income of the household. We would have ideally liked to include information on contraception use at different points of the woman's life, but that data is unavailable and also its inclusion could cause endogeneity bias. So we instead use an indicator variable, EVERUSE, if the woman ever used contraception. This in some sense captures the woman's attitude towards (and/or awareness of) family planning and choosing the duration between children rather that leaving it "in the hands of God".

The World Bank has emphasized the role of household income (or expenditure) on malnutrition and child mortality (Behrman and Knowles, 1999). This is because household income or expenditure reflects household command over different inputs, e.g., food, clothing, residence, sanitation, medical care, in the child health production function. The demographic and health surveys (of which NFHS is a part) do not collect information on household income/expenditure. We instead compute and include in the set of explanatory variables a composite asset index (PCASSET). This is a composite asset index and we use principal component analysis to construct this index from household ownership of agricultural land, farm equipment, cycle, scooter, car, radio and television. In the child mortality equation we include a number of household level infrastructural variables: main source of drinking and non drinking water and the main source of toilet. These capture the environment in which the child is born and could have a significant effect on child health.

For the Indian sample we also include several religion dummies HINDU, MUSLIM and OTHERS (including other minority groups Christians, Buddhists and Jains). The omitted category here is the Sikhs (the majority religion in the state).[11] In the Pakistani sample there are no non-Muslim households. Importance of religion on family formation is well documented in the literature. It has been argued that Muslim societies are often predisposed to high fertility and child mortality (compared to non-Muslims). While some argue that this is related to lack of women's autonomy in decisions regarding fertility and child health as promoted by Islam (Basu, 1992), empirical tests do not always support this (see for example Morgan, Stash, Smith and Mason, 2002). Thus our choice of samples offers an excellent opportunity to understand the distinctive demographic trends in Muslim and non-Muslim communities in these two neighbouring states in at least two ways. First no one can deny that religious identity is intertwined with socio-economic status, health infrastructure and other unobservable determinants of mortality. Second, the welfare state can effectively intervene to assist demographic development, as has been experienced elsewhere in the developing world (e.g., China). While the sample households are socio-culturally very similar because of their common origin, they were partitioned in 1947 primarily on the basis of their religion and have been ruled by very different types of institutions (religious/political and interaction between the two, if any) since then. The latter could shape fertility and mortality differently in the two provinces either directly in terms of contraception use and/or indirectly through differential female literacy rates.

[11] We had experimented with various religious categories and in most cases religion dummies were not significant. Following one referee's suggestion, we have in the final specification considered Sikhs as the reference category to make it more uniform, though the result did not change much from other specifications we tried.

In addition to the parental/household characteristics we should ideally include supply side factors that affect child health and duration between children. Data on such supply side variables is however not available. So we include a rural residence dummy to capture the effects of all omitted supply side variables.[12]

In the hazard of child mortality regression, we include indicators for the type of toilet and the main source of water. These capture the environment in which the child is born and grows up and could have a statistically significant effect on the health of the child.

Finally we include birth cohort dummies in both the regressions. These birth cohort dummies, to some extent, capture the over all trends in these demographic indicators and reflect the relative importance of the demand and supply factors. In the Indian sample we include three dummies: the child was born between 1970 and 1980 (YEARB2), born between 1980 and 1990 (YEARB3) and born after 1990 (YEARB4). The reference category is that the child was born before 1970. For Pakistan we include one indicator variable: born after 1977. This year is an important one in the history of Pakistan as Zia-ul-Huq assumed power in this year and embarked on a process of full scale Islamisation of the country, its programs and policies.

The baseline hazards are specified as splines. The two baseline hazards $T_1(t)$ and $T_2(t)$ measure the duration dependence of survival and subsequent birth. These essentially measure the time varying risk of child mortality and subsequent child birth from the time the child is at risk of the event. The time dependency starts once the child

[12] Note however that this is the residence during the time of the survey and may not always reflect the residence during the first five year's of a child's life (assuming marriage-related migration between rural and urban areas for women). Indeed this is also true for some other variables including household assets, access to toilet or safe drinking water. Thus the effects of some of these variables on mortality may not be pronounced in our analysis.

is born. Several specifications of the baseline hazard were tried and ultimately we selected the specification that fitted the data best: for the Pakistani sample, we chose four nodes at 12, 18, 24 and 30 months to characterise the baseline hazard in the log hazard of duration to next birth equation; the corresponding nodes for the Indian sample were 12 and 24 months. Similarly, we use different nodes to characterise the baseline hazard in the log hazard of child mortality equation in the two samples: it was 1 month for the Pakistani sample while 3 nodes 3, 6 and 9 months were used for the Indian sample.

4. Regression Results:

For each country we estimate a system of correlated hazards. We compute and present separate estimates for (1) the first born and only child and (2) the non-first born. There are two reasons for separating the sample in this manner. First in many developing countries the first born child is treated very differently, compared to the latter born children. The second is a more pedantic reason: for the first born, by definition sibling composition of elder siblings and mortality status of elder siblings is not defined. Both these variables are however interesting in their own right. Our regression results actually justify this form of sample stratification. In both samples we exclude the youngest child of the woman who is sterilized at the time of the survey since the question of spacing is no longer relevant for them.

Results corresponding to a number of different specifications are presented. *Specification* 1 presents the single equation (uncorrelated) estimates of the log hazard of duration to next birth or child mortality, ignoring unobserved mother level heterogeneity. *Specification* 2 presents the single equation estimates of the log hazard of duration to next birth and child mortality, but this time we account for unobserved mother level heterogeneity. *Specification* 3 presents the correlated hazard estimates from the joint estimation of the log hazard of duration to next birth and the log hazard of child mortality. Finally *Specification* 4 presents the single equation estimates for the first born. Since here we have one observation per child, there is no unobserved mother level heterogeneity to account for.

Tables 2 and 3 respectively present the full set of results for India and Pakistan respectively. The estimates for the unobserved heterogeneity components (σ_n^2, σ_s^2 and ρ) show that ignoring unobserved mother level heterogeneity results in biased estimates and second single equation estimates are inconsistent: the correlation between the unobserved heterogeneity coefficients (ρ) is statistically significant in both regressions (though only weakly so for the Indian sample). Note however that the sign of the correlation coefficient is different in the two samples: while the unobserved heterogeneity that increases birth interval is associated with a lower mortality (negative coefficient) in India, it is exactly the opposite (positive coefficient) in Pakistan. The preferred specification (for the non-first born children) is therefore given by *Specification* 3 and we will discuss these results and highlight the differences compared to the other two specifications.

A negative (positive) and statistically significant coefficient associated with any particular variable in the log hazard of duration to next birth regression implies that this variable reduces (increases) the hazard of having a subsequent child and increases (reduces) the duration between the index child and the next child. Similarly a negative (positive) and statistically significant coefficient associated with any particular variable in the log hazard of child mortality regression implies that this variable reduces (increases) the hazard of child mortality and increases (decreases) the number of days the child was alive.

Birth Spacing:

The coefficient estimates and associated standard errors for the log hazard of duration to next birth are presented in Tables 2 and 3 for India and Pakistan respectively. In each case we discuss the results for the non-first born (*Specification 3*) and the first born children (*Specification* 4).

We start with a discussion of the results for the *non-first born children in India*. Clearly hazard of having a subsequent sibling is significantly lower after the first two years of a child's life. There is a very strong child replacement effect: an increase in the child mortality (measured by an increase in the number of months the child is alive) has a negative and statistically significant effect on the log hazard of time to next birth i.e. on the duration to next birth. There is, however, no evidence of son preference – the gender of the index child does not have a statistically significant effect on the log hazard of duration to next birth. Surprisingly, the log hazard of duration to next birth is significantly higher (equivalently the duration to next birth is significantly lower) following the birth of twins. One possibility is that it reflects a couple's desire to complete family formation (to attain target family size) within a given period rather than updating family planning in view of the birth of a twin.

The log hazard of duration to next birth is significantly lower for older women (more than 25). This is a rather surprising result since one would expect older mothers to have lower duration between births (if they have not reached their desired family size), given the biological constraints on child bearing. What could explain this result is the fact that in many developing countries, most women get married and have children early and children born to women above 25 are few and far between.

Not surprisingly the log hazard of duration to next birth is significantly lower (and the duration to next birth significantly higher) for educated women, relative to women who are illiterate. What is however interesting is that there is evidence of a threshold level of education that must be attained by the woman before her educational attainment starts having a statistically significant effect on the duration between births. Educated women are better able to understand the benefits (both to the mother and to the child) of increased duration following birth. It appears that women need to have sufficient education before this works successfully. Father's education however does not have a statistically significant effect on the log hazard of duration to next birth. The log hazard of duration to next birth is significantly lower when the mother has ever used contraceptives. Contraceptive awareness is therefore related to the desire to increase the duration between successive births though in the absence of more detailed data on contraceptive use we cannot really elaborate on this argument. None of the religion variables is however significant (reference category Sikhs). Finally, two other results are worth noting: the log hazard of duration to next birth is significantly higher for rural residents (possibly because of the generally lower levels of literacy and/or poor awareness/conservative attitude towards modern contraception) as it is for births in the 1990s (no declining trend is noted here).

Next, considering the *non-first born children in Pakistan*, results appear to be somewhat different to those obtained for India. There is no evidence of the child replacement effect: child mortality (measured by an increase in the number of months the child is alive) does not have a statistically significant effect on the log hazard of duration to next birth i.e. on the duration to next birth. If anything, the coefficient estimate of NEXT is actually positive, though nowhere close to being even weakly statistically significant. We do however find significant evidence of son preference: the log hazard of duration to next birth is significantly lower (or the duration to next birth is significantly higher) following the birth of a son. In addition we also find that parents prefer to have at least one son: the log hazard of duration to next birth is significantly higher (equivalently the duration to next birth is significantly lower) if all existing children are girls. Finally the hazard of subsequent birth is significantly lower is any of the previous children born to the woman have died.

As in the Indian case, the mother's age at the time of birth has a statistically significant effect on the log hazard of duration to next birth. The log hazard of duration to next birth is significantly lower (equivalently the duration to next birth higher) for mothers' aged 21 and higher relative to mothers' aged 20 or below. In addition it is worth noting that there is a monotonic relationship between the age of the mother and the duration between successive births: the older the mother at the time of the birth of the index child, the longer is the duration between the index child and the next. The increased duration to the next birth for older mothers in this case is essentially a reflection of the fact that women get married and have children fairly early. Not surprisingly the log hazard of duration to next birth is significantly lower (and the duration to next birth significantly higher) for educated women, relative to women who are illiterate. Further the higher the educational attainment of the mother, the stronger is the effect of mother's educational attainment on the duration to next birth. Interestingly father's educational attainment does not have a statistically significant effect on the log hazard of duration to next birth. The log hazard of duration to next birth is significantly lower when the mother has ever used contraceptives. Once again we could think of contraceptive use as being indicative of the desire to increase the duration between successive births, but in the absence of more detailed data on contraceptive use we cannot really elaborate on this argument.

The composite asset index is negative and statistically significant implying that the log hazard of duration to next birth is significantly lower for children born to wealthy parents. One possible explanation could be that wealthy households are more likely to be more educated (often a close correlation between education and wealth is observed in these low-income economies) and thus more aware of the potential benefits of increased duration between successive births. The log hazard of duration to next birth is significantly lower for rural residents – this again is quite a surprising result. The log hazard of duration to next birth is significantly higher for children born after 1977. It appears that the Islamisation of the country and the reduced emphasis on family planning and maternal health programs that happened after 1977 had a strong and adverse effect on the hazard of birth spacing.

The results for the *first born* are quite different to those for the non-first born and this holds for both samples. Notice that possibly due to the much smaller sample size and exclusion of unobserved mother-specific heterogeneity (and therefore these uncorrelated estimates are not corrected for the simultaneity bias) these coefficient estimates are less precise for the first-born (only one observation per mother and that is why we cannot include unobserved heterogeneity). Nevertheless, several interesting differences (vis-à-vis non-firstborn children) are worth noting. *For Pakistani sample*, there is no evidence of gender bias – the boy dummy is not statistically significant. There is however significant evidence of the child replacement effect. With the exception of contraceptive use none of the parental, household and community characteristics have a statistically significant effect on the log hazard of duration to the second birth. Now turning to the *Indian case*, there is evidence of significant wealth effect in that children from more wealthy families have lower hazard of having a younger sibling.

Child Survival:

We now turn to a discussion of the regression results for the log hazard of child mortality. The coefficient estimates and associated standard errors are presented in Tables 2 and 3 respectively for India and Pakistan. Once again in each case we discuss the results for the non-first born (*Specification 3*) and then the first born children (*Specification* 4).

First of all, hazard of child mortality is significantly lower after six months of a child's life among *non-first born children in India*. There is evidence of significant resource constraint/sibling competition effect: an increase in the duration between child i and child $i+1$ significantly reduces the log hazard of mortality of child i (equivalently increases the life of the child). Although there is no significant gender difference in child mortality, there is some evidence of pro-male bias among resource constrained parents. In particular, mortality hazard of the index child is significantly lower if all the previous children are female. Mortality of older siblings is associated with a significant increase in the log hazard of mortality of the index child. Not surprisingly the hazard of child mortality is significantly higher if the child is a part of a twin birth.

With the exception of the asset index none of the parental/household characteristics have a statistically significant effect on the log hazard of child mortality. The log hazard of child mortality is significantly lower for wealthier households. We also find that access to modern toilet significantly (though only weakly) lowers the mortality hazard of the sample children; significance of this variable, at least in part, reflects the role of provision of services (supply side factors) in improving child mortality. Finally relative to children born before 1970, the log hazard of child mortality is significantly lower for children born during the period 1980 – 1990, though this declining trend is not observed in the 1990s.

As with the Indian case, there is significant evidence of resource constraint/sibling rivalry effect in the *Pakistani sample*: an increase in the duration between child i and child $i+1$ significantly reduces the log hazard of mortality of child i (equivalently increases the life of the child).[13] Mortality of older siblings is associated with a significant increase in the log hazard of mortality of the index child.

Not surprisingly the log hazard of child mortality is significantly lower for children with educated mothers and interestingly the log hazard of child mortality is significantly lower for children born to older mothers. Father's educational attainment does not have a statistically significant effect on the log hazard of child mortality. The log hazard of child mortality is significantly higher for children born in rural households – possibly reflecting the poorer health services and facilities that are available in the rural areas. Finally we find that the log hazard of child mortality is significantly higher for children born after 1977. Again it appears that the Islamisation of the country and

---

[13] In addition, an increase in prior birth spacing (the duration between child $i-1$ and child i) significantly reduces the log hazard of mortality of child i (see Table A2).

the reduced emphasis on family planning and maternal health programs that happened after 1977 had a strong and adverse effect on child health.

The results for the *first born* are again quite interesting in their own right. First even for the first child there is strong evidence of resource constraint/sibling rivalry effect in Pakistan. Second we find that there is a threshold level of education that must be attained by the mother (more than completed primary schooling) before maternal education starts having an effect on the log hazard of child mortality for the first child. Again the log hazard of child mortality is significantly higher for children born after 1977. The central result still holds for the first born Indian children in that longer spacing significantly lowers mortality hazard. Interesting difference can also be noted for the Indian sample in this respect. While the male dummy is insignificant for the non-first born children, first born male children have significantly higher mortality hazards. Although there is no significant wealth effect in this context (unlike the non-first born children), access to both modern toilet and safe drinking water plays a significant role to improve child survival for the first-born.

One other issue is worth noting here. An anonymous referee has noted that the effect of resource constraints on the index child might be different depending not only on the gender of the current child but also on the gender composition of the elder children. To examine this issue we interact the gender of the child (BOY) with the ALLPREVFEM dummy and included it as an additional explanatory variable in both regressions. In this case the non-interacted coefficient of ALLPREVFEM gives us the effect on girls and the interaction coefficient gives us the difference effect. These results are not presented but are available on request. While the introduction of this interaction term made no difference in results for the Indian sample, it made one important difference in the mortality equation of the Pakistani sample. In particular, the hazard of mortality (of the index child) is significantly higher if all children are girls and the difference estimate is negative and statistically significant. It appears that a male child is significantly better off (possibly in terms of resources devoted to him leading to better health outcomes) if all the elder siblings are girls. This is not unexpected given that our analysis takes account of the correlation between fertility and mortality decisions (see discussion in section 3, pp. 13).

Finally, we compare and contrast the Indian and the Pakistani results with respect to household decisions in birth spacing and child survival. This brings out some interesting similarities and differences between the two states divided by the partition in 1947 on the basis of religion. One clarification is however noteworthy here. There are no Non-Muslim households in the Pakistan sample while Muslims are the minority in the Indian state (we include several religion dummies HINDU, MUSLIM and OTHERS and the omitted category here is the Sikhs). The religion dummies are however not statistically significant so that we cannot directly identify any religion specific effects either in spacing or in mortality within India, though the differences in determinants of fertility and mortality decisions by households in the two provinces would indirectly reflect differences in religious and political institutions, shaping the state policies towards education, employment and family planning/welfare, especially since the introduction of the Islamic state in Pakistan after 1977; however it is difficult to isolate these effects directly in our sample.

<u>Differences in household behaviour in India and Pakistan</u>

1. While there is a very strong child replacement effect on fertility in India, the

corresponding effect is not statistically significant in Pakistan.

2. Mother's educational attainment has very strong effects on the duration between births (as well as for duration of survival) in Pakistan, but less so in India. Any level of mother's literacy lowers the hazard of subsequent birth in Pakistan[14] however, more than primary schooling of the mother is required to have any perceptibly favourable effect on subsequent childbirth (effect of mother's education is however statistically not significant for child survival) in India.
3. There is some evidence of son preference in both samples, though the nature is somewhat different. Thus the duration to the next birth in Pakistan is significantly lower following the birth of a girl and also if all the previous children are also girls. In contrast, the mortality risk of the current child in India is significantly lower if all the previous children are girls. If however we include a gender interaction term with all previous children being female, the effect turns out to be significant only in Pakistan so that boys enjoy a lower mortality hazard if all previous siblings are girls.
4. Finally, compared to the pre-1977 period, the hazard of subsequent birth is significantly higher in Pakistan while the trend has been just opposite in India. The latter seems to highlight the role of active population programme initiated by the government of India since early 1950s; in contrast there was no official population programme in Pakistan until early 1990s.

Thus one major factor that appears to drive the differences in results between Indian and Pakistani Punjabi provinces is the effect of the mother's educational attainment on birth spacing and child survival, other things remaining unchanged. The average parental literacy, especially, mother's literacy is significantly higher in India.

[14] We are unable to identify if the lower female literacy among Pakistani households is a product of Islamic beliefs since there are no non-Muslim households in the Pakistani sample.

Thus, a marginal increase in parental literacy would have a less pronounced effect in the Indian Punjab (compared to the Pakistani Punjab). The differential nature of son preference between the two provinces is also quite interesting and not highlighted in previous studies generally based on single equation estimates of child health functions. We believe this is one advantage of the correlated hazard model used in our analysis that takes account of the correlation between spacing and child mortality, generally ignored in the literature.

While we cannot isolate the effects of religion, state policy and their interaction, if any, in our analysis, differences in the results from these two adjoining provinces tend to highlight the significance of religion and state policy on household demographic behaviour.

5. Effect of Breastfeeding

While our analysis has made a strong case for an inverse relationship between birth spacing and child mortality in both Indian and Pakistani Punjab, we have not so far discussed any possible biological/behavioural mechanism affecting this relationship. As the anonymous referees have pointed out, breastfeeding might play an important role in this respect: not only is the duration of breastfeeding closely correlated with birth interval, but also it improves the likelihood of survival among infants. First the primary link between breastfeeding and birth spacing arises because breastfeeding increases the *post partum amenorrhoea*, i.e., the time between a birth and resumption of the menstruation. Secondly, breast milk is extremely nutritious for the infant and also contains immunological elements that provide protection against different forms of infections among infants; thus breastfeeding improves the survival chances of infants. Given this close biological link between breastfeeding on the one hand and birth spacing and survival on the other, we will in this section examine the effects of breastfeeding on birth spacing and child mortality in our samples.

The major obstacle to include the possibility of breastfeeding in our model arises from the fact that the data on the duration on breastfeeding was not collected for the full sample. It was only collected for children born during the period 1986 – 1991 (the five years preceding the survey) in Pakistan and during the period 1989 – 1992 (the three years preceding the survey) in India. So the sample size is now significantly smaller. In particular we now have information on 2153 children born to 1328 mothers in Pakistan and 1613 children born to 1135 mothers in India. For the Pakistan sample, the average duration of breastfeeding is nearly 14 months and for the Indian sample the corresponding duration is 13 months (for the sample of children that have ever been breastfed). It is worth noting that 8.6% of the children in Pakistan and 4.53% of the children in India have never been breastfed.

To examine whether breastfeeding has any effect on the duration to next birth (NEXT) and child mortality (SURV) we present in Figures 1 and 2 the effect of ever breastfeeding on NEXT and SURV respectively (for Pakistan and India). Figure 1 implies that the hazard of having a younger sibling is not significantly different depending on whether or not the child has been breastfed. Indeed using a non-parametric Wilcoxon test we are not able to reject the null hypothesis of equality of the survivor functions ($\chi^2(1) = 1.33; p-value = 0.2493$ for Pakistan and $\chi^2(1) = 0.00; p-value = 0.9916$ for India). On the other hand, Figure 2 implies that for both Pakistan and India, the hazard of child mortality is significantly higher for infants that are never breastfed and this difference is particularly significant for the first 6 months of the child's life. Not surprisingly using a non-parametric Wilcoxon test, we reject the null hypothesis of equality of the survivor functions ($\chi^2(1) = 996.41; p-value = 0.0000$ for Pakistan and $\chi^2(1) = 307.11; p-value = 0.0000$ for India).

However, merely looking at whether or not the child has ever been breastfed or not does not tell us the full story and indeed there is a fair amount of variation in the duration of time a child has been breastfed. For example for the Pakistan sample, 50.3% of the sample children have been breastfed for 12 months or less (this includes children that are being breastfed at the time of the survey) while 9% of the sample children have been breastfed for more than 2 years. For the Indian sample, 56.17% of the sample children have been breastfed for 12 months (again including children that are still breastfed at the time of the survey) and more than 9% of the children have been breastfed for more than 2 years. Figure 3 shows the smoothed hazard estimates of spacing and duration of breastfeeding for the sample of children that have ever been breastfed and including children that are being breastfed at the time of the survey) for Pakistan and India. These hazard estimates for spacing and breastfeeding are strikingly similar and in both samples: there is therefore evidence of a close correspondence between spacing and duration of breastfeeding in both samples. Accordingly we argue that breastfeeding is a biological mechanism of spacing in both our samples and one could be used as an instrument for the other. This is further confirmed in the empirical analysis that we undertake below.

Introduction of breastfeeding however complicates our estimation strategy. Mothers who choose to breastfeed their children (and/or choose to breastfeed their children longer) are not necessarily a random subset of all mothers. On the one hand these could be women with a strong preference for healthy children (lower child mortality) and/or a longer duration between children and thus want to reduce the impact of the resource constraint and the sibling-rivalry effects through longer breastfeeding (this is favourable self-selection). On the other hand these women may have some private information about their own health and, are therefore particularly concerned with the health of their children. Given this adverse self-selection, they may choose to breastfeed their children. Thus ignoring adverse self selection could lead to an under estimation of the effect of breastfeeding, while ignoring favourable self selection actually may cause the effects of breastfeeding on birth outcomes to be overstated. In other words the duration of breastfeeding could be endogenous in both the birth spacing and the child mortality regressions. In an attempt to obtain the selectivity-corrected effect of breastfeeding on birth spacing and child mortality we therefore jointly estimate (the duration of) breastfeeding with the duration to next birth and child mortality hazard regressions, after allowing for cross-correlations between breastfeeding, birth spacing and child mortality. This requires us to add a third equation to the system of equations (1) and (2). The log hazard of the duration (in months) child i born to woman j was breastfed is given by:

$$h_{ij}^{b}(t) = \phi_0 + \phi_1 T_3(t) + \phi_2 X_{3ij} + \lambda_j^b + \varepsilon_{ij}^b$$

where X_{3ij} denotes a set of exogenous explanatory variables that affect the duration of breastfeeding. As before, the unexplained component of the log hazard of the duration of breastfeeding is broken up into a component that is purely random $\left(\varepsilon_{ij}^b\right)$ and a component that is common to all children born to the same mother $\left(\lambda_j^b\right)$, capturing the mother level unobserved heterogeneity. Again the duration of breastfeeding (DURBF) is censored if a child is being still breastfed at the time of the survey. The mother specific unobserved heterogeneity in the duration of breastfeeding equation could be correlated with the mother specific unobserved components of the error terms in the spacing and survival equations, so

$$\begin{pmatrix} \lambda^b \\ \lambda^n \\ \lambda^s \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_b^2 & \sigma_{bn} & \sigma_{bs} \\ \sigma_{bn} & \sigma_n^2 & \sigma_{ns} \\ \sigma_{bs} & \sigma_{ns} & \sigma_s^2 \end{pmatrix} \right)$$

As before, the full specification model is estimated jointly using Full Information Maximum Likelihood (FIML) method.

There is one other data issue that needs attention. Remember that data availability restricts us to using a considerably smaller sample when we want to account for breastfeeding. We are therefore unable to run separate estimates for the first born and the non-first born children. Regressions are run for “all children”. Clearly the variables ALLPREVM and ANYPREVD are not defined for the first born. In this case we code them as zero. The problem is that if we restrict ourselves to the non-first born we are left with only 1747 children for Pakistan and 1136 children for India and the maximum likelihood estimate fails to converge when we jointly estimate DURBF, NEXT and SURV. This exercise proves to be particularly difficult for India where breastfeeding information is available only for the children born in the last 3 years of survey (as opposed to 5 years in the Pakistani case). As a result, there is not enough variation in the mother-specific heterogeneity in the Indian sample – in about 64% cases there was only one child born in the last three years of the survey and we fail to obtain the correlated hazard estimate for India including breastfeeding. Perhaps this difference also highlights the differential pattern of childbearing in the two neighbouring provinces. As a result, we here highlight the results obtained from the Pakistani sample.

When we estimate this three-equations system (including duration of breastfeeding), we fail to generate meaningful correlated hazard estimates in that we cannot reject the null hypothesis of zero correlation between the unobserved components of the error terms in the log hazard of duration of breastfeeding equation and the log hazard of duration to next birth equations (ρ_{bn}) and the unobserved components of the error terms in the log hazard of duration of breastfeeding equation and the log hazard of child survival equations (ρ_{bs}). As an alternative we go back to the original two equations systems to jointly determine hazard of next birth and child mortality, with duration of breastfeeding (DURBF) as an additional (exogenous) explanatory variable. We estimate and present four sets of results in Table 4. *Specification 1* presents the uncorrelated hazard estimates of $h_{ij}^{n}(t)$ and $h_{ij}^{s}(t)$ when SURV and DURBF are both included as explanatory variables in the log hazard of time to next birth $\left(h_{ij}^{n}(t)\right)$ equation and when NEXT and DURBF are both included as explanatory variables in the log hazard of child mortality $\left(h_{ij}^{s}(t)\right)$ equation. *Specification 2* presents the corresponding correlated hazard estimates. While not surprisingly, an increased duration of breastfeeding reduces the hazard of next birth and also reduces the hazard of child mortality, the coefficient estimates of SURV and NEXT in the two regressions are very different to what one would expect. In particular, an increase in the duration between births increases the hazard of child mortality, while the longer the child is alive the greater is the hazard of next birth. The question now is: how do we explain these rather surprising results. One could perhaps attribute it to the bias generated by the close correspondence between NEXT and DURBF that we discussed earlier.[15] This conjecture is further confirmed when we include either NEXT or DURBF as in specification (3) or (4). In S*pecification 3* we do not include NEXT as an explanatory variable in the log hazard of child mortality regression, but we do include DURBF suggesting that DURBF is a proxy for NEXT in this regression. In *Specification 4* we do not include DURBF as an additional explanatory variable but do include NEXT. In either case we get back our central result that after controlling for everything else, greater spacing or greater duration of breastfeeding is likely to lower mortality. In other words, this analysis not only strengthens our earlier result (inverse relationship between spacing and mortality), but also enables us to empirically establish that breastfeeding is a useful instrument for spacing in our samples.

6. Conclusion

This paper examines the two-way relationship between birth interval and child survival and compares the behaviour of households in the Indian and Pakistani provinces of Punjab. Birth interval and child survival are modelled here as correlated hazard processes to address the bias generated by the simultaneity between spacing and survival, allowing for mother-specific unobserved heterogeneity. We find evidence of significant mutual dependence between birth interval and child survival in both samples. There is also evidence of a close correspondence between duration of spacing and breastfeeding and accordingly we argue that spacing is a good instrument of duration of breastfeeding in our samples.

[15] The correlation between DURBF and NEXT is 0.5545 (significantly different from zero).

Our analysis also enables us to differentiate the demographic behaviour of predominantly Muslim and non-Muslim households in the Indian and Pakistani Punjab provinces who share a common socio-cultural background until they were partitioned in 1947 on religious ground. First, we identify significant differences especially with respect to effects of female literacy and son preference on spacing and mortality in the two provinces that remain much unexplored in earlier studies. We believe this is an advantage of our superior methodology that takes account of the correlation between spacing and mortality and also allows us to correct for the possible simultaneity bias between spacing and child survival. Finally there is evidence that the hazard of subsequent birth has been declining in India in recent decades though the trend has been just opposite in Pakistan, especially since the introduction of the Islamic state after 1977. While a part of these differences could be explained by differences in religion, lower literacy (especially female literacy) and the interaction between the two, a part has to be attributed to rather passive official population policy in Islamic Pakistan (compared to secular India) for much of the post-independence period.

REFERENCES

Arnold, F., M. K. Choe and T. K. Roy (1998): "Son Preference, the Family Building Process and Child Mortality in Bangladesh", *Population Studies*, **52**, 301 – 315.

Basu, A. (1992). Culture, the Status of Women and Demographic Behaviour, Oxford: Clarendon Press.

Becker, G. S. (1981). A Treatise on the Family. Cambridge, Harvard University Press.

Becker, G. S. and H. G. Lewis (1973): "On the Interaction between Quantity and Quality of Children", *Journal of Political Economy*, **81**(2(Part2)), S279 - S288.

Behrman, J. R. and J.C. Knowles (1999): "Household Income and Child Schooling in Vietnam", *The World Bank Economic Review*, 13(2) pp. 211 - 56.

Behrman, J. R. and B. L. Wolfe (1984): "Evidence on Nutrition Demand", *Journal of Development Economics*, 105 - 128.

Brien, M. J. and L. Lillard, A. (1994): "Education, Marriage, and First Conception in Malaysia", *Journal of Human Resources*, **29**(4), 1166 - 1204.

Brien, M. J., L. A. Lillard and L. J. Waite (1999): "Interrelated Family-Building Behaviors: Cohabitation, Marriage and Nonmarital Conception", *Demography*, **36**(4), 535 - 551.

Caldwell, J. A. (1986): "Routes to Low Mortality in Poor Countries", *Population and Development Review*, **12**, 171 - 220.

Chamberlain, G. and Z. Grilliches (1975): "Unobservables with a Variance Component Structure: Ability, Schooling and the Economic Success of Brothers", *International Economic Review*, **16**(2), pp. 422 - 49.

Cleland, J. and Z. A. Sathar (1984): "The Effect of Birth Spacing on Childhood Mortality in Pakistan", *Population Studies*, **38**(4), 401 - 418.

Dasgupta, M. (1987): "Selective Discrimination against Female Children in Rural Punjab", *Population and Development Review*, **13**.

Gangadharan, L. and P. Maitra (2003): "The Effect of Education on the Timing of Marriage and First Birth in Pakistan", *Journal of Quantitative Economics*, **New Series, 1**(1), 114 - 133.

Garg, A. and J. Morduch (1998): "Sibling Rivalry and the Gender Gap: Evidence from Child Health Outcomes in Ghana", *Journal of Population Economics*, **11**(4), 471 - 493.

Gribble, J.N. (1993): "Birth intervals, gestational age and low birth weight: are the relationships confounded?", *Population Studies*, **47**, 133-146.

Horton, S. (1988): "Birth Order and Child Nutrition Status: Evidence from the Philippines", *Economic Development & Cultural Change*, **36**(2), 341 - 354.

Hussain, R.F., F. Firkee and H. W. Berendes (2000): "The Role of Son Preference in Reproductive Behaviour in Pakistan", Bulletin of World Health Organisation', 78(3) pp. 379-88.

Jejeebhoy, S. J. and Z. A. Sathar (2001): "Women's Autonomy in India and Pakistan: The Influence of Religion and Region", *Population and Development Review*, **27**(4), 687 - 712.

Lillard, L. A. (1993): "Simultaneous Equations for Hazards: Marriage Duration and Fertility Timing", *Journal of Econometrics*, **56**(1-2), 189 - 217.

Maitra, P. and S. Pal (2004): "Early Childbirth, Health Inputs and Child Mortality: Recent Evidence from Bangladesh", *Mimeo, Monash University*.

Morgan, S. P., S. Stash, H. L. Smith and K. O. Mason (2002): "Muslim and non-Muslim Differences in Female Autonomy and Fertility: Evidence from Four Asian Countries", *Population and Development Review*, **28**(3), 515 - 537.

Murthi, M., A. Guio and J. Dreze (1995): "Mortality, Fertility, and Gender Bias in India: A District-Level Analysis", *Population and Development Review*, **21**(4), 745 - 782.

Muthrayappa, R., M. K. Choe, F. Arnold and T. K. Roy (1997): "Son Preference and Its Effect on Fertility in India", *National Family Health Survey Subject Reports No. 3*.

Pal, S. (1999): "An Analysis of Childhood Malnutrition in Rural India: Role of Gender, Income and Other Household Characteristics", *World Development*, **27**(7), 1151 - 1171.

Panis, C. W. A. (1994): "The Piecewise Linear Spline Transformation with an Application to Age at First Marriage", *Stata Technical Bulletin*, **18**, 27 - 29.

Rosenzweig, M. R. and T. P. Schultz (1982): "Market Opportunities, Genetic Endowments and Intrafamily Resource Distribution: Child Survival in Rural India", *American Economic Review*, **72**(4), 803-815.

Sathar, Zeba A.; Casterline, John B (1998) 'The Onset of Fertility Transition in Pakistan' *Population and Development Review,*. 24(4).773-796.

Table 1. Means and standard deviations of Selected Demographic Variables

Variable	India	Pakistan
Oldest Child	0.3102	0.2218
	(0.45)	(0.42)
Youngest Child	0.2709	0.2098
	(0.44)	(0.41)
Dead at the Time of the Survey	0.0826	0.1338
	(0.28)	(0.34)
SURV (in years, sample not censored)	11.52	13.95
	(17.88)	(26.28)
NEXT (in years, sample not censored)	30.26	27.90
	(16.47)	(16.06)
Children ever born	4.02	4.22
	(1.65)	(3.00)
Highest School Attainment of Mother: Primary School (EDUCM1)	0.1639	0.1171
	(0.37)	(0.32)
Highest School Attainment of Mother: More than Primary School	0.20	0.1908
(EDUCM2)	(0.40)	(0.39)
Highest School Attainment of Father: Primary School (EDUCF1)	-	0.1488
		(0.36)
Highest School Attainment of Father: More than Primary School	-	0.4353
(EDUCF2)		(0.50)
If father is literate (LITDAD)	0.5827	-
	(0.49)	

Note: Standard deviations in parentheses.

Table 2: Spacing and Survival Hazard Estimates for Full Sample, India

	Non-first born children						First born children	
	Uncorrelated Hazard: With/Without Unobserved Heterogeneity				Correlated Hazard		Uncorrelated Hazard: Without Unobserved Heterogeneity	
	Spacing		Survival		Spacing	Survival	Spacing	Survival
	Without Heterogeneity	With Heterogeneity	Without Heterogeneity	With Heterogeneity				
Splines for hazard of subsequent birth								
0-12 months	0.3355 ***	0.3348 ***			0.3331 ***		0.8077 ***	
	(0.0176)	(0.0177)			(0.0183)		(0.0825)	
12-24 months	0.1202 ***	0.1263 ***			0.1263 ***		0.0882 ***	
	(0.0046)	(0.0048)			(0.0049)		(0.006)	
>24 months	-0.0034 **	0.0021			0.002		-0.0157 ***	
	(0.0014)	(0.0018)			(0.002)		(0.0014)	
Splines for hazard of child mortality								
0-3 months			1.2107 ***	1.2062 ***		1.2026 ***		1.371 ***
			(0.1122)	(0.1132)		(0.114)		(0.161)
3-6 months			0.232	0.2296		0.2292		0.0591
			(0.1446)	(0.145)		(0.1458)		(0.2457)
6-9 months			-0.1871 *	-0.1852 *		-0.1845 *		-0.3533 *
			(0.0994)	(0.0994)		(0.1004)		(0.1941)
>9 months			-0.0300 ***	-0.0301 ***		-0.0303 ***		0.0067
			(0.0042)	(0.0042)		(0.0044)		(0.0061)
CONSTANT	-7.615 ***	-7.6263 ***	-0.838	-0.8788	-7.5252 ***	-0.5319	-12.989 ***	0.0854
	(0.2326)	(0.2424)	(0.9217)	(1.009)	(0.2545)	(0.9003)	(0.9718)	(0.7007)
Child-specific variables								
MALE	-0.0222	-0.0075	-0.074	-0.0634	-0.014	-0.0795	-0.0668	0.3600 **
	(0.0333)	(0.0356)	(0.1014)	(0.1026)	(0.0368)	(0.1045)	(0.0425)	(0.144)
NEXT			-0.0186 ***	-0.0185 ***		-0.0241 ***		-0.031 ***

			(0.0031)	(0.0031)		(0.0066)		(0.0051)
SURV	-0.007 ***	-0.0082 ***			-0.0098 ***		-0.0084 ***	
	(0.0008)	(0.0009)			(0.0018)		(0.0013)	
TWIN	0.5387 ***	0.7517 ***	1.3371 ***	1.4137 ***	0.7498 ***	1.3673 ***	1.4492***	0.9905 **
	(0.0776)	(0.0826)	(0.2146)	(0.2333)	(0.0855)	(0.2599)	(0.4965)	(0.4648)
ANYPREVD	0.03	0.0241	0.9556 ***	0.8817 ***	0.0807	0.9072 ***		
	(0.0476)	(0.0521)	(0.112)	(0.1368)	(0.0612)	(0.1388)		
ALLPREVFEM	-0.0207	-0.0142	-0.4438 ***	-0.4463 ***	-0.0134	-0.4539 ***		
	(0.0362)	(0.0409)	(0.133)	(0.1346)	(0.041)	(0.1367)		
AGEM2	-0.0225	-0.0367	-0.1745	-0.1818	-0.0288	-0.1568	-0.0393	0.0683
	(0.0791)	(0.0851)	(0.2177)	(0.2214)	(0.0858)	(0.2223)	(0.049)	(0.159)
AGEM3	-0.1379 *	-0.1661 *	0.1032	-0.1192	-0.1560 *	-0.0748	-0.2421 ***	0.1262
	(0.0802)	(0.0876)	(0.2241)	(0.2292)	(0.0884)	(0.2307)	(0.075)	(0.2503)
AGEM4	-0.1974 **	-0.2132 **	-0.3814	-0.4055	-0.2037 **	-0.3848	-1.0961 ***	0.1187
	(0.0921)	(0.1007)	(0.2698)	(0.2773)	(0.1017)	(0.2786)	(0.2047)	(0.802)
Household-specific variables								
EDUCM1	-0.0058	-0.0078	0.0092	0.0066	-0.0058	0.008	0.0121	-0.0464
	(0.0506)	(0.0594)	(0.1658)	(0.1732)	(0.06)	(0.175)	(0.0619)	(0.1999)
EDUCM2	-0.1346 **	-0.1410 *	0.3206	0.3389	-0.1438 **	0.3579	-0.0867	-0.352
	(0.0634)	(0.0726)	(0.2135)	(0.2202)	(0.0732)	(0.2211)	(0.0695)	(0.2455)
LITDAD	-0.0383	-0.0414	0.1338	0.1345	-0.04	0.1351	0.0318	-0.1642
	(0.0403)	(0.0485)	(0.1232)	(0.1294)	(0.0489)	(0.1294)	(0.0544)	(0.1659)
HINDU	-0.0103	-0.0075	0.0243	0.0282	-0.009	0.0324	0.0187	0.0817
	(0.0346)	(0.0419)	(0.1148)	(0.1224)	(0.0421)	(0.1232)	(0.046)	(0.1492)
MUSLIM	-0.0036	-0.0305	-0.3446	-0.3313	-0.0325	-0.3083	0.5875 ***	0.1595
	(0.115)	(0.1409)	(0.3586)	(0.3762)	(0.1425)	(0.3871)	(0.1885)	(0.4661)
OTHERS	0.0251	-0.028	0.3881	0.3636	-0.0003	0.4159	-0.2203	-0.9424
	(0.1556)	(0.1931)	(0.4808)	(0.5069)	(0.204)	(0.5274)	(0.173)	(1.0334)
EVERUSE	-0.0869 **	-0.0992 **			-0.1023 **		0.1569 ***	
	(0.0353)	(0.042)			(0.0424)		(0.0462)	
PCASSET	-0.0061	-0.0068	-0.3736 ***	-0.3754 ***	-0.0014	-0.3825 ***	-0.0489 *	-0.019
	(0.022)	(0.0262)	(0.0734)	(0.0768)	(0.0266)	(0.0769)	(0.0255)	(0.0931)

MODERN TOILET			-0.2519	-0.2466		-0.2462*		-0.403 **
			(0.1553)	(0.162)		(0.131)		(0.2004)
SAFE WATER			-1.0945	-1.127		-1.2756		-1.293 **
			(0.8839)	(0.9766)		(0.8355)		(0.5897)
RURAL	0.1083 ***	0.1189 **	0.076	0.0965	0.1141 **	0.0696	-0.0054	-0.2565
	(0.0409)	(0.0488)	(0.1556)	(0.1639)	(0.0492)	(0.1647)	(0.051)	(0.1944)
YEARB2	-0.0764	-0.0829	-0.3701 *	-0.3762 *	-0.079	-0.3784 *	0.1145	-0.3082
	(0.0748)	(0.0843)	(0.1997)	(0.2054)	(0.0865)	(0.2146)	(0.0755)	(0.218)
YEARB3	-0.0082	-0.0085	-0.4628 **	-0.4594 **	-0.0025	-0.4829 **	0.0907	-0.534 **
	(0.0743)	(0.0861)	(0.2037)	(0.2127)	(0.0876)	(0.2188)	(0.0746)	(0.2228)
YEARB4	0.5245 ***	0.5296 ***	-0.3836	-0.3916	-0.4965 ***	-0.4644	-0.0504	-0.562 **
	(0.1158)	(0.1278)	(0.2966)	(0.313)	(0.1332)	(0.3214)	(0.1165)	(0.2794)
Unobserved Heterogeneity								
σ_v^2		0.3250 ***				0.3205 ***		
		(0.0395)				(0.0415)		
σ_σ^2				0.3566 **		0.3418 *		
				(0.1773)		(0.1926)		
ρ						-0.9789 *		
						(0.4724)		

NOTE: Asymptotic standard errors in parentheses; Significance: '*'=10%; '**'=5%; '***'=1%.

Table 3. Spacing and survival hazard estimates for Full Sample, Pakistan

	Non First Born						First Born	
	Uncorrelated Hazard: With/Without Unobserved Heterogeneity				Correlated Hazard		Uncorrelated Hazard: Without Unobserved Heterogeneity	
	Spacing		Survival		Spacing	Survival	Spacing	Survival
	Without Heterogeneity	With Heterogeneity	Without Heterogeneity	With Heterogeneity				
Splines for Hazard of Subsequent Birth								
0 – 12 months	0.6028 ***	0.6049 ***			0.6050 ***		0.6276 ***	
	(0.0293)	(0.0293)			(0.0298)		(0.0491)	
12 – 18 months	-0.1032 ***	-0.1003 ***			-0.0991 ***		-0.0971 ***	
	(0.0115)	(0.0115)			(0.0117)		(0.0201)	
18 – 24 months	0.2211 ***	0.2240 ***			0.2244 ***		0.1977 ***	
	(0.0101)	(0.0102)			(0.0103)		(0.0183)	
24 – 30 months	-0.0631 ***	-0.0540 ***			-0.0540 ***		-0.0334 **	
	(0.0077)	(0.0079)			(0.0079)		(0.0149)	
> 30 months	-0.0283 ***	-0.0269 ***			-0.0267 ***		-0.0221 ***	
	(0.0014)	(0.0014)			(0.0014)		(0.0027)	
Splines for Hazard of Child Mortality								
0 – 1 month			-0.8113 *	-0.7027		-0.7414		-1.0705
			(0.4813)	(0.4905)		(0.4928)		(0.9036)
> 1 month			-0.0616 ***	-0.0624 ***		-0.0618 ***		-0.0593 ***
			(0.0014)	(0.0014)		(0.0014)		(0.0023)
CONSTANT	-10.2254 ***	-10.2861 ***	-3.3949 ***	-3.5496 ***	-10.2727 ***	-3.6349 ***	-10.2101 ***	-3.2692 ***
	(0.3388)	(0.3414)	(0.4803)	(0.4969)	(0.3487)	(0.4982)	(0.5636)	(0.8768)
Child Specific Variables								
BOY	-0.0560 **	-0.0676 **	-0.0375	-0.0324	-0.0696 **	-0.0429	-0.0476	0.2041 *
	(0.0282)	(0.0292)	(0.0669)	(0.0680)	(0.0293)	(0.0689)	(0.0505)	(0.1210)
SURV	-0.0005 **	-0.0004			0.0002		-0.0009 ***	
	(0.0002)	(0.0002)			(0.0003)		(0.0003)	
NEXT			-0.0100 ***	-0.0097 ***		-0.0051 ***		-0.0138 ***

			(0.0012)	(0.0013)		(0.0017)		(0.0027)
ALLPREVFEM	0.0547	0.0918 **	0.0498	0.0353	0.0786 **	0.0795		
	(0.0334)	(0.0381)	(0.0811)	(0.0876)	(0.0385)	(0.0893)		
ANYPREVD	0.0466 *	0.0073	0.9277 ***	0.8162 ***	-0.0896 **	0.7107 ***		
	(0.0279)	(0.0343)	(0.0670)	(0.0834)	(0.0382)	(0.0864)		
AGEM2	-0.1513 ***	-0.1797 ***	-0.3164 ***	-0.3276 ***	-0.1826 ***	-0.3776 ***	0.0186	-0.3721 **
	(0.0435)	(0.0461)	(0.0836)	(0.0887)	(0.0471)	(0.0911)	(0.0567)	(0.1445)
AGEM3	-0.3376 ***	-0.3991 ***	-0.5797 ***	-0.6160 ***	-0.3978 ***	-0.7212 ***	-0.1139	0.0815
	(0.0454)	(0.0503)	(0.0991)	(0.1072)	(0.0510)	(0.1097)	(0.1010)	(0.2217)
AGEM4	-0.6309 ***	-0.7403 ***	-1.0579 ***	-1.1151 ***	-0.7426 ***	-1.3550 ***	-0.2272	-0.1242
	(0.0870)	(0.0997)	(0.1740)	(0.1983)	(0.0999)	(0.2066)	(0.2168)	(0.5210)
AGEF2	0.1435 **	0.1286	-0.4042 ***	-0.4339 ***	0.1114	-0.5082 ***	0.1211	-0.2576
	(0.0720)	(0.0799)	(0.1368)	(0.1551)	(0.0795)	(0.1579)	(0.0848)	(0.1894)
AGEF3	0.1702 ***	0.2062 ***	-0.0221	-0.0200	0.2037 ***	0.0354	0.0476	-0.0586
	(0.0444)	(0.0485)	(0.0889)	(0.0957)	(0.0491)	(0.0969)	(0.0735)	(0.1789)
AGEF4	-0.0243	-0.0240	0.0366	0.0358	-0.0299	0.0300	0.0421	0.0939
	(0.0353)	(0.0407)	(0.0845)	(0.0964)	(0.0415)	(0.1004)	(0.0635)	(0.1587)
Household Specific Variables								
EDUCM1	-0.1241 **	-0.1340 *	-0.3169 **	-0.3399 **	-0.1489 **	-0.3313 **	0.0590	-0.1170
	(0.0544)	(0.0686)	(0.1371)	(0.1617)	(0.0702)	(0.1683)	(0.0848)	(0.2053)
EDUCM2	-0.1896 ***	-0.2049 ***	-0.4273 ***	-0.4544 ***	-0.2147 ***	-0.5267 ***	0.0297	-0.7066 ***
	(0.0471)	(0.0595)	(0.1495)	(0.1717)	(0.0612)	(0.1841)	(0.0889)	(0.2569)
EDUCF1	-0.0017	0.0053	0.0762	0.0632	0.0064	0.0677	-0.0376	0.2826 *
	(0.0371)	(0.0493)	(0.0870)	(0.1143)	(0.0509)	(0.1234)	(0.0809)	(0.1609)
EDUCF2	-0.0452	-0.0514	-0.0145	-0.0142	-0.0586	-0.0409	-0.1099 *	-0.1677
	(0.0345)	(0.0448)	(0.0924)	(0.1130)	(0.0462)	(0.1180)	(0.0652)	(0.1584)
RURAL2	-0.1453 ***	-0.1487 ***	0.4155 ***	0.4181 ***	-0.1531 ***	0.3661 ***	-0.0709	0.0911
	(0.0394)	(0.0519)	(0.0989)	(0.1247)	(0.0531)	(0.1301)	(0.0756)	(0.2078)
PCASSET	-0.1046 ***	-0.1168 ***	-0.1486 *	-0.1464	-0.1251 ***	-0.1537	-0.0277	0.0058
	(0.0250)	(0.0330)	(0.0890)	(0.1109)	(0.0338)	(0.1137)	(0.0494)	(0.1596)
YEAR_B77	0.1490 ***	0.1589 ***	0.2346 ***	0.2593 ***	0.1128 **	0.2801 ***	0.0295	0.3961 ***
	(0.0382)	(0.0415)	(0.0733)	(0.0837)	(0.0467)	(0.0866)	(0.0690)	(0.1396)
EVERUSE	-0.1405 ***	-0.1631 ***			-0.1733 ***		-0.3262 ***	
	(0.0308)	(0.0409)			(0.0411)		(0.0618)	
NO TOILET			-0.2684 **	-0.2724 *		-0.2015		0.1978

		(0.1048)	(0.1391)	(0.1455)	(0.2363)
PIPED DRINKING WATER		0.0312	0.0197	-0.0129	0.6577
		(0.2374)	(0.2802)	(0.2856)	(0.4186)
PIPED OTHER WATER		-0.0386	-0.0056	0.0253	-0.7199 *
		(0.2394)	(0.2819)	(0.2858)	(0.4243)
Unobserved Heterogeneity					
σ_n^2	0.3039 ***			0.3288 ***	
	(0.0300)			(0.0306)	
σ_s^2			0.5506 ***	0.6400 ***	
			(0.0705)	(0.0754)	
ρ				0.8943 ***	
				(0.1220)	

NOTE: Asymptotic standard errors in parentheses;
Significance: '*'=10%; '**'=5%; '***'=1%.

Table 4. Effect of Breastfeeding: Spacing and Survival Hazard Estimates
(for Children Born in the last 5 years in Pakistan)

	Specification 1 Uncorrelated hazard		Specification 2 Correlated hazard		Specification 3 Correlated hazard		Specification 4 Correlated hazard	
	SURV and DURBF included in the duration hazard equation	NEXT and DURBF included in the mortality hazard equation	SURV and DURBF included in the duration hazard equation	NEXT and DURBF included in the mortality hazard equation	SURV not included in the duration hazard equation	NEXT not included in the mortality hazard equation	DURBF not included in the duration hazard equation	DURBF not included in the mortality hazard equation
	Spacing	Survival	Spacing	Survival	Spacing	Survival	Spacing	Survival
Splines for Hazard of Subsequent Birth								
0 – 12 months	0.7458 *** (0.0873)		0.7543 *** (0.0903)		0.7494 *** (0.0900)		0.7323 *** (0.0883)	
12 – 18 months	-0.0115 (0.0304)		-0.0049 (0.0315)		-0.0146 (0.0310)		-0.0411 (0.0308)	
18 – 24 months	0.2232 *** (0.0273)		0.2338 *** (0.0281)		0.2219 *** (0.0277)		0.1993 *** (0.0272)	
24 – 30 months	0.0123 (0.0243)		0.0244 (0.0248)		0.0070 (0.0244)		-0.0183 (0.0235)	
> 30 months	0.0363 *** (0.0114)		0.0461 *** (0.0122)		0.0359 *** (0.0112)		0.0215 * (0.0110)	
Splines for Hazard of Child Mortality								
0 – 1 month		7.3975 (7.7975)		8.2618 (8.9906)		8.5887 (9.3661)		1.6759 (2.0875)
> 1 month		-0.1038 *** (0.0234)		-0.1007 *** (0.0226)		-0.0882 *** (0.0234)		-0.1818 *** (0.0144)
CONSTANT	-12.1181 *** (1.0010)	-11.2494 (7.7328)	-12.5333 *** (1.0461)	-12.7486 (8.9895)	-11.9035 *** (1.0282)	-11.4037 (9.3126)	-12.1855 *** (1.0179)	-6.1034 *** (1.9702)
Child Specific Variables								
BOY	0.0767 (0.0848)	0.5575 *** (0.1982)	0.0746 (0.0931)	0.3489 (0.2507)	0.0599 (0.0824)	0.6032 *** (0.2098)	0.0695 (0.0762)	0.3819 ** (0.1628)
SURV	0.0110 ***		0.0286 ***				-0.0004	

	(0.0033)		(0.0041)				(0.0034)	
NEXT		0.0230 **		0.1040 ***				-0.0114
		(0.0092)		(0.0139)				(0.0087)
ALLPREVFEM	-0.1075	-0.2732	-0.2167	-0.4187	-0.0687	-0.1818	-0.0916	-0.3219
	(0.1281)	(0.2958)	(0.1348)	(0.3694)	(0.1222)	(0.3175)	(0.1093)	(0.2638)
ANYPREVD	-0.0970	0.0816	-0.2634 **	0.0027	-0.1502	-0.0046	-0.1854 *	0.2025
	(0.1177)	(0.2169)	(0.1256)	(0.2685)	(0.1136)	(0.2301)	(0.1002)	(0.1946)
DURBF	-0.0748 ***	-0.4443 ***	-0.0971 ***	-0.5019 ***	-0.0634 ***	-0.4731 ***		
	(0.0064)	(0.0293)	(0.0074)	(0.0307)	(0.0055)	(0.0304)		
AGEM2	-0.2280	0.1498	-0.1958	0.3644	-0.2329 *	0.1450	-0.1940	0.1184
	(0.1465)	(0.3120)	(0.1526)	(0.3585)	(0.1403)	(0.3274)	(0.1235)	(0.2693)
AGEM3	-0.5859 ***	0.4623	-0.4521 **	0.3810	-0.6066 ***	0.7023	-0.5098 ***	0.0485
	(0.1682)	(0.4212)	(0.1774)	(0.4945)	(0.1652)	(0.4523)	(0.1434)	(0.3433)
AGEM4	-1.0007 ***	-1.0308 **	-1.0925 ***	-1.6509 **	-0.9204 ***	-1.2532 **	-0.9126 ***	-1.4258 ***
	(0.2165)	(0.4703)	(0.2390)	(0.6673)	(0.2076)	(0.5433)	(0.1871)	(0.4502)
AGEF2	-0.1188	-0.3697	-0.2351	-0.9656 **	-0.0463	-0.6938 *	-0.1010	-0.7943 ***
	(0.1677)	(0.3480)	(0.1885)	(0.4744)	(0.1590)	(0.3658)	(0.1367)	(0.2855)
AGEF3	0.0575	0.4456	-0.0190	0.3060	0.0766	0.5441 *	0.0827	0.1554
	(0.1132)	(0.2892)	(0.1260)	(0.3285)	(0.1094)	(0.3026)	(0.0990)	(0.2324)
AGEF4	-0.1128	0.4995	-0.1151	0.4064	-0.0799	0.6038	-0.0364	0.4785
	(0.1279)	(0.3691)	(0.1434)	(0.4737)	(0.1241)	(0.4099)	(0.1143)	(0.3066)
Household Specific Variables								
EDUCM1	-0.0342	-0.1624	-0.0368	-0.4404	-0.0384	-0.3579	0.0466	-0.4036
	(0.1578)	(0.4301)	(0.1727)	(0.5055)	(0.1517)	(0.4250)	(0.1312)	(0.3448)
EDUCM2	-0.1283	-0.9393 **	-0.1762	-1.5530 ***	-0.1094	-1.4289 ***	0.0631	-0.6533 *
	(0.1631)	(0.4643)	(0.1803)	(0.5137)	(0.1570)	(0.4662)	(0.1392)	(0.3687)
EDUCF1	0.2440 *	0.7563 **	0.2678	0.6898 *	0.2330 *	0.6780 **	-0.1777	0.1005
	(0.1436)	(0.3062)	(0.1675)	(0.4131)	(0.1379)	(0.3439)	(0.1135)	(0.2653)
EDUCF2	0.2947 **	-0.2863	0.2642 *	-0.2482	0.2775 **	-0.5296	0.2522	-0.4163
	(0.1264)	(0.3216)	(0.1388)	(0.3786)	(0.1205)	(0.3291)	(0.0000)	(0.2545)
RURAL2	-0.0594	0.7550 *	0.0004	0.6388	-0.0677	0.4613	-0.0812	0.6069 *
	(0.1460)	(0.4224)	(0.1636)	(0.4787)	(0.1422)	(0.4442)	(0.1188)	(0.3133)
PCASSET	-0.1254	-0.3964	-0.1051	-0.1976	-0.1138	-0.5150	0.0005	0.2771
	(0.0922)	(0.3185)	(0.1014)	(0.3592)	(0.0902)	(0.3332)	(0.0742)	(0.2520)
EVERUSE	-0.0021		-0.0468		-0.0312		-0.0908	

	(0.1143)	(0.1240)	(0.1091)	(0.0965)
NO TOILET	-0.3639	-0.1710	-0.6374	-0.2803
	(0.4349)	(0.5264)	(0.4699)	(0.3225)
PIPED DRINKING WATER	0.9875	1.0419	0.9149	1.3453 **
	(0.7289)	(0.8842)	(0.6392)	(0.6467)
PIPED OTHER WATER	-0.1969	-0.3111	-0.2188	-0.7941
	(0.7649)	(0.8689)	(0.6594)	(0.6508)
Unobserved Heterogeneity				
σ_n^2	0.8718 ***	1.1332 ***	0.7484 ***	0.5237 ***
	(0.1083)	(0.0990)	(0.1034)	(0.1165)
σ_s^2	2.1228 ***	2.5128 ***	2.3016 ***	1.2599 ***
	(0.1970)	(0.2402)	(0.2175)	(0.1809)
ρ		0.8573 ***	0.2224 ***	0.8115 ***
		(0.0470)	(0.0714)	(0.2275)

NOTE: Asymptotic standard errors in parentheses;
Significance: '*'=10%; '**'=5%; '***'=1%

Figure 1: The Effect of Breastfeeding on the Hazard of Duration to Next Birth (NEXT)

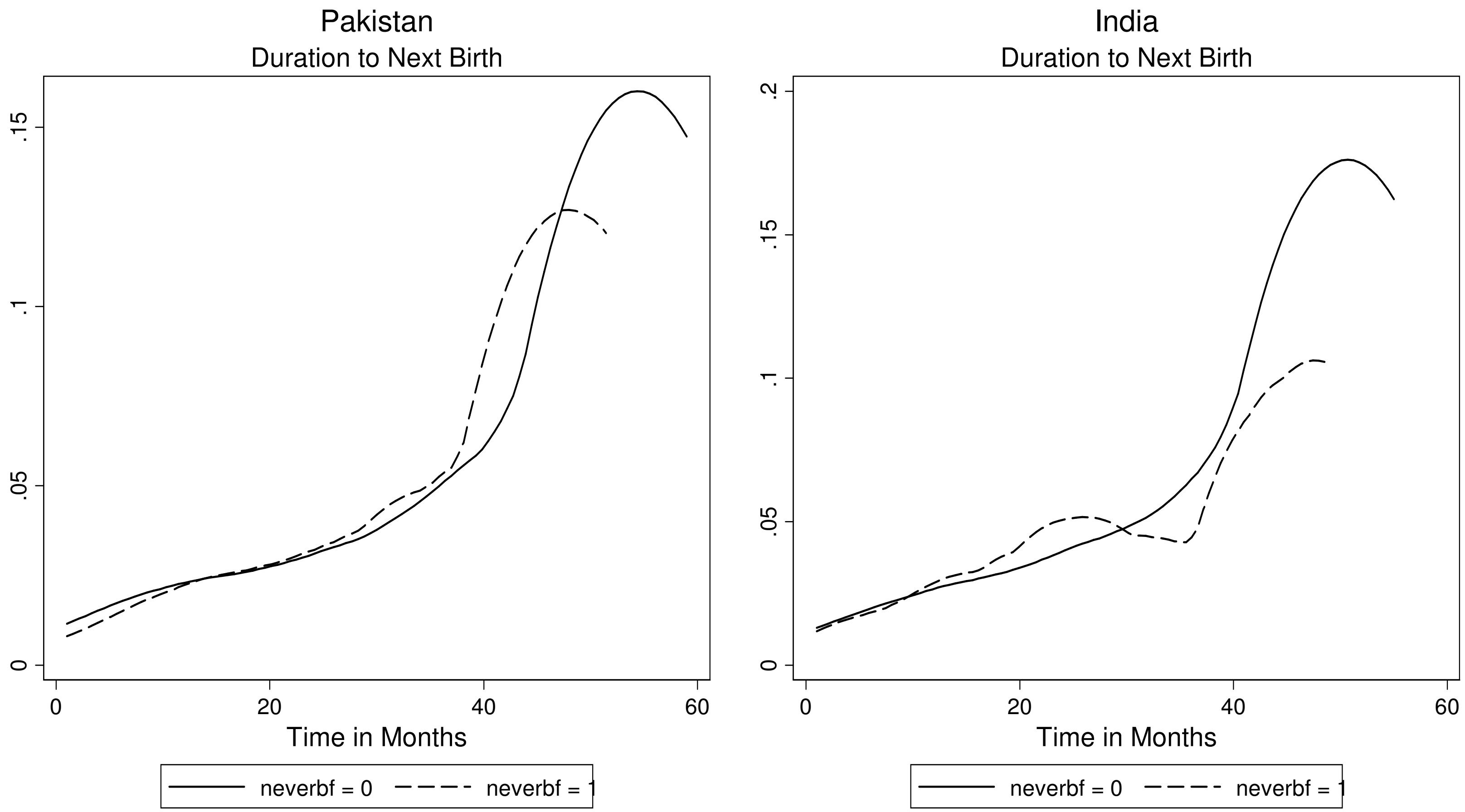


Figure 2: The Effect of Breastfeeding on the Hazard of Child Survival (SURV)

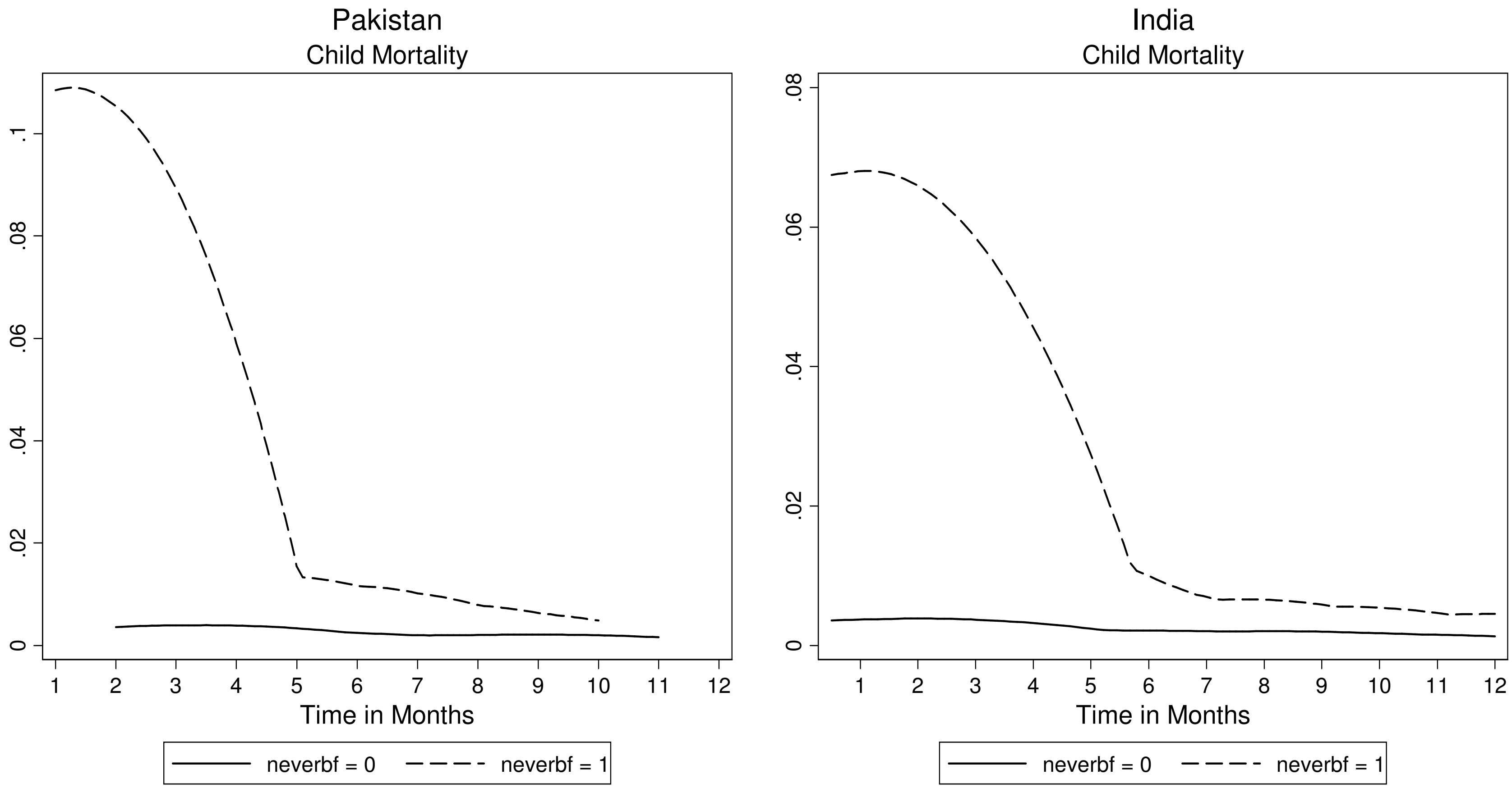


Figure 3. Smoothed Hazard Estimates for India

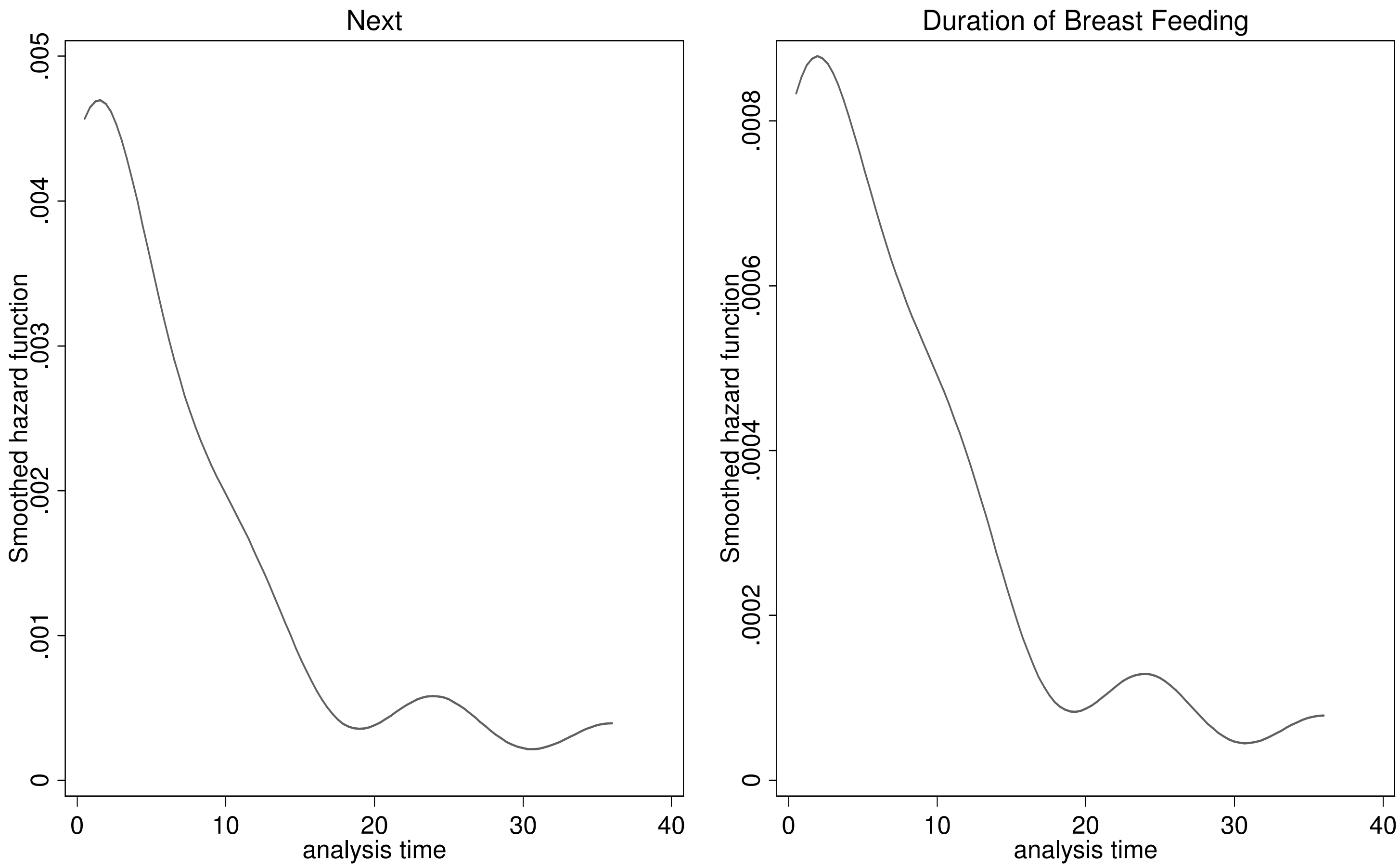
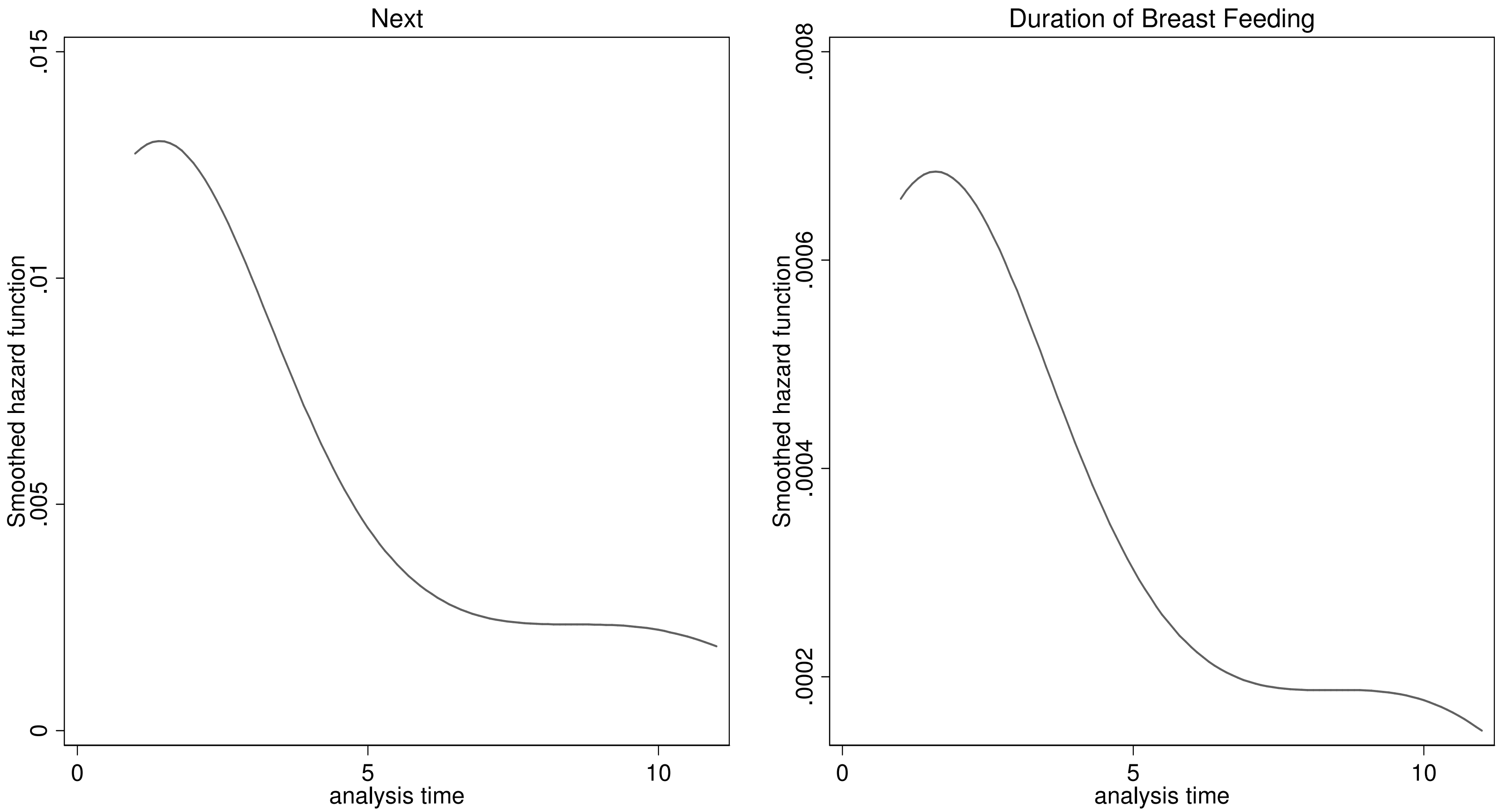


Figure 3. Smoothed Hazard Estimates for Pakistan (continued)



APPENDIX

Table A1: Definition of regression variables

Variable	India	Pakistan
Parental Characteristics		
AGEF1		= 1 if the father is less than 26 years old at the time of birth (Reference Category).
AGEF2		= 1 if the father is aged 26 – 30 at the time of birth
AGEF3		= 1 if the father is aged 31 – 35 at the time of birth.
AGEF4		= 1 if the father is aged more than 36 at the time of birth.
AGEM1	= 1 if the mother is less than 21 years old at the time of birth. (Reference Category).	= 1 if the mother is less than 21 years old at the time of birth. (Reference Category).
AGEM2	= 1 if the mother is aged 21 – 25 at the time of birth.	= 1 if the mother is aged 21 – 25 at the time of birth.
AGEM3	= 1 if the mother is aged 26 – 30 at the time of birth.	= 1 if the mother is aged 26 – 30 at the time of birth.
AGEM4	= 1 if the mother is aged more than 30 at the time of birth.	= 1 if the mother is aged more than 30 at the time of birth.
EDUCM0	= 1 if mother is illiterate (Reference Category).	= 1 if mother is illiterate (Reference Category).
EDUCM1	= 1 if highest education attained by mother is primary school.	= 1 if highest education attained by mother is primary school.
EDUCM2	= 1 if highest education attained by mother is more than primary school.	= 1 if highest education attained by mother is more than primary school.
EDUCF0		= 1 if father is illiterate (Reference Category).
EDUCF1		= 1 if highest education attained by father is primary school.
EDUCF2		= 1 if highest education attained by father is more than primary school.
LITDAD	=1 if the father is literate (can read write).	
EVERUSE	=1 if ever used contraception.	=1 if ever used contraception.
Child and Sibling Characteristics		
BOY		= 1 if child is a boy.
POLDF	Proportion of Elder siblings that are females. Not defined for first child.	Proportion of Elder siblings that are females. Not defined for first child.
ALLPREVFEM	= 1 if all children born to the mother are girls. Not defined for first child.	= 1 if all children born to the mother are girls. Not defined for first child.
ANYPREVD	= 1 if any child born to the mother earlier has died.	= 1 if any child born to the mother earlier has died.
SURV	Months lived before dying. Equals Age if alive at time of survey (censored).	Months lived before dying. Equals Age if alive at time of survey (censored).
NEXT	Duration (in months) between two successive children. Equals Age if last child (censored).	Duration (in months) between two successive children. Equals Age if last child (censored).
PRIOR	Prior duration between two successive children. Not defined for first child.	Prior duration between two successive children. Not defined for first child.
YEARB2	= 1 if the child is born between 1970 and 1980.	
YEARB3	= 1 if the child is born between 1980 and 1990.	

YEARB4	= 1 if the child is born after 1990.	
YEARB_77		= 1 if born before 1977
Other Household Characteristics		
HINDU	= 1 if comes from a Hindu Family.	
SIKH	= 1 if comes from a Sikh Family.	
PCASSET	Composite Indicator of Assets.	Composite Indicator of Assets.
SAFEWATER	=1 if have access to safe drinking water.	
MODTOILT	= 1 if have access to modern toilet.	
NO TOILET		= 1 if household does not have a toilet
PIPED DRINKING WATER		= 1 if the main source of drinking water is piped.
PIPED OTHER WATER		= 1 if the main source of other water is piped.
Community Characteristics		
RURAL	= 1 if household resides in a rural area.	= 1 if household resides in a rural area.

Table A2: Estimates of Child Mortality Hazard. Including PRIOR. Pakistan only.

	Uncorrelated Hazard without Unobserved Heterogeneity	**Uncorrelated Hazard with Unobserved Heterogeneity**	**Correlated Hazard**
Splines for Hazard of Child Mortality			
0 – 1 month	-0.8146 *	-0.6826	-0.7170
	(0.4820)	(0.4957)	(0.5058)
> 1 month	-0.0629 ***	-0.0627 ***	-0.0622 ***
	(0.0014)	(0.0014)	(0.0015)
CONSTANT	-2.8496 ***	-2.9919 ***	-3.0716 ***
	(0.4834)	(0.5044)	(0.5109)
Child Specific Variables			
BOY	-0.0421	-0.0419	-0.0475
	(0.0668)	(0.0682)	(0.0698)
NEXT	-0.0083 ***	-0.0086 ***	-0.0065 ***
	(0.0013)	(0.0013)	(0.0017)
PRIOR	-0.0277 ***	-0.0287 ***	-0.0262 ***
	(0.0025)	(0.0025)	(0.0026)
ALLPREVFEM	0.0748	0.0494	0.0772
	(0.0819)	(0.0883)	(0.0902)
ANYPREVD	0.8812 ***	0.7611 ***	0.7196 ***
	(0.0665)	(0.0837)	(0.0863)
AGEM2	-0.2251 ***	-0.2160 **	-0.2444 ***
	(0.0844)	(0.0902)	(0.0930)
AGEM3	-0.4238 ***	-0.4373 ***	-0.5000 ***
	(0.1019)	(0.1115)	(0.1151)
AGEM4	-0.8502 ***	-0.8824 ***	-1.0049 ***
	(0.1747)	(0.1987)	(0.2088)
AGEF2	-0.4169 ***	-0.4498 ***	-0.4772 ***
	(0.1371)	(0.1545)	(0.1565)
AGEF3	-0.0432	-0.0462	-0.0216
	(0.0895)	(0.0968)	(0.0979)
AGEF4	0.0335	0.0346	0.0290
	(0.0854)	(0.0968)	(0.1000)
Household Specific Variables			
EDUCM2	-0.3104 **	-0.3365 **	-0.3343 **
	(0.1368)	(0.1622)	(0.1652)
EDUCM3	-0.4571 ***	-0.4891 ***	-0.5111 ***
	(0.1505)	(0.1735)	(0.1819)
EDUCF2	0.1004	0.0901	0.0953
	(0.0881)	(0.1159)	(0.1203)
EDUCF3	-0.0029	-0.0081	-0.0137
	(0.0923)	(0.1124)	(0.1155)
RURAL2	0.4481 ***	0.4623 ***	0.4329 ***
	(0.0987)	(0.1242)	(0.1285)
PCASSET	-0.1399	-0.1459	-0.1545
	(0.0889)	(0.1108)	(0.1142)
YEAR_B77	0.2396 ***	0.2698 ***	0.2787 ***
	(0.0737)	(0.0842)	(0.0871)
NO TOILET	-0.2634 **	-0.2830 **	-0.2466 *
	(0.1050)	(0.1393)	(0.1457)
PIPED DRINKING	0.0068	0.0014	-0.0159
WATER	(0.2347)	(0.2796)	(0.2833)
PIPED OTHER	-0.0082	0.0117	0.0155
WATER	(0.2363)	(0.2805)	(0.2836)
Unobserved Heterogeneity			
σ_s^2		0.5566 ***	0.5819 ***

		(0.0726)	(0.0799)
ρ			0.9620 *** (0.3621)

NOTE: Asymptotic standard errors in parentheses;
Significance: '*'=10%; '**'=5%; '***'=1%.